

COMPUTATIONAL STUDY OF THE DEAMINATION
REACTION OF CYTOSINE

MANSOUR HUSSEIN ALMATARNEH





Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-33468-3

Our file Notre référence

ISBN: 978-0-494-33468-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Computational Study of the Deamination Reaction of Cytosine

By

Mansour Hussein Almatarneh

A thesis

**Submitted to the School of Graduate Studies in Partial
Fulfilment of the Requirements for the Degree of
Doctor of Philosophy**

**Department of Chemistry
Memorial University of Newfoundland**

September, 2007

St. John's,

Newfoundland

Canada

ABSTRACT

The decomposition reaction of formamidine yielding hydrogen cyanide and ammonia has been studied first as a simple model for the intramolecular and intermolecular hydrogen rearrangement of cytosine. The gas phase decomposition of formamidine predicted a high activation energy of 259 kJ mol^{-1} at the G3 level of theory. Adding one and two water molecules catalysed the reaction by forming a cyclic hydrogen bonded transition state, reducing the barrier to 169 and 151 kJ mol^{-1} at the G3 level, respectively. The PCM solvent model predicts a significant lowering of the free energy of activation.

The mechanism for the deamination reaction of cytosine with H_2O , OH^- , and $\text{H}_2\text{O}/\text{OH}^-$ to produce uracil was investigated using *ab initio* (HF and MP2) levels and B3LYP DFT calculations. All pathways in the cytosine deamination produce an initial tetrahedral intermediate followed by several conformational changes. The final intermediate for all pathways dissociates to products via a 1–3 proton shift. Two pathways for the deamination reaction of cytosine with H_2O and OH^- were found. The activation energy for the rate-determining steps of deamination of cytosine with H_2O for pathways A and B are 221 and 260 kJ mol^{-1} at the G3MP2 level of theory, respectively. The deamination of cytosine with H_2O by either pathway is therefore unlikely because of the high barriers involved. Deamination

with OH^- through pathway C resulted in the lowest activation energy, 203 kJ mol^{-1} at the G3MP2 level of theory.

The deamination with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$ in which the water molecules acted as a solvent and a catalyst was also investigated. Seven pathways for the deamination reaction for these systems were found. We found that the barrier for the water-mediated 1–3 proton shift is reduced by 46 kJ mol^{-1} at the G3MP2 level of theory. We also found that the addition of the second water molecule reduces the barriers for both rate-determining steps by 31 kJ mol^{-1} . The activation energy for the rate-determining step, the formation of tetrahedral intermediate for pathway D is 115 kJ mol^{-1} at the G3MP2 level of theory, in excellent agreement with the experimental value. This work shows, for the first time, a plausible mechanism for the deamination of cytosine and accounts for the observed experimental activation energy ($117 \pm 4 \text{ kJ mol}^{-1}$).

Acknowledgments

In the name of God, the Most Gracious, the Most Merciful, I thank God (Allah) first for helping me finish this program and achieve my dreams and indeed throughout my life. There are several individuals I would like to thank for their guidance and support while this work was being completed.

First of all, I would like to greatly acknowledge my supervisor, Professor Raymond Poirier, the best advisor and teacher I could have wished for, for all the guidance, encouragement, advice, and support he has offered me during my study under his supervision. He has had a great influence on my career and I owe much of my interest in chemistry to him. I can not thank him enough for his invaluable assistance in the completion of my degree. Indeed without his guidance, this thesis would not have been possible.

I would like to acknowledge Dr. Christopher Flinn for his guidance, advice and support throughout the research projects, as well as his effort in proofreading the manuscripts. I especially appreciate his fine touch in reviewing the first chapter in this thesis. I would also like to thank my supervisory committee, Dr. Robert Davis and Dr. Erika Merschrod, for their help and suggestions.

I would like to thank our entire research group, as well as Prof. Paul Mezey's research group, for their continual support, help and encouragement throughout my program. In the two offices in which I have worked, I was surrounded by knowledgeable and friendly

colleagues, Tammy Gosse, Aisha El-Sherbiny, Joshua Hollett, Mohammad Islam, Csaba Szakacs, Eva Simon, and Kushan Saputantri. I am especially grateful to Josh for his effort in proofreading the second chapter in this thesis. I am also fortunate to have the help and support of many great friends in St. John's. I really have to thank Nicole Edwards and Terri Ryan for their help and support. I am also grateful for the friendship of Melanie Russell and Kevin Pomroy, my first connection to Newfoundland lifestyle. I would like to thank the Chemistry Department, the School of Graduate Studies, and NSERC for financial support during my program. I would also like to thank ACEnet for computer time.

Last but not least, I am forever indebted to my loving parents, Amneh and Hussein, who have been a great source of strength through all my life. I would like to acknowledge my sisters, Nawarah, Jamelih, Fatima, Kadejeh, Fadwa, Nehaih, and my brothers Khaled and Naser for their unconditional support, encouragement and love. I would like especially to thank my oldest brother, Khaled (Abu Rabie), for standing by me through good and bad times. Rabie and Hammam were also extremely important to me while I was writing this thesis. I would like to extend a special thanks to: my brothers' in law, especially Abu Nabawi; my nieces; my nephews; and all my friends in Jordan. Their continued support and encouragement throughout the completion of my degree was greatly appreciated.

Thank you all!

Dedications

The Lord has blessed me with a wonderful family to whom
this thesis is dedicated.

~ To my dear mother and the memory of my beloved father ~

TABLE OF CONTENTS

Abstract	ii
Acknowledgments	iv
Dedication	vi
Table of Contents	vii
List of Tables	xii
List of Figures	xv
List of Schemes	xx
List of Abbreviations and Symbols	xxi
List of Publications From This Thesis	xxv
Chapter 1 General Introduction	
1.1 Overall Goals and Objectives	1
1.2 The Nucleic Acid (DNA Molecule)	5
1.2.1 Nucleic Acid bases	6
1.2.2 Hydrogen Bonding: Base Pairing	7
1.2.3 Nucleoside and Nucleotide	8
1.3 Cytosine	10
1.3.1 Chemical and Physical Properties	10

1.3.2 Tautomerism of Cytosine	14
1.4 Uracil	16
1.4.1 Chemical and Physical Properties	16
1.4.2 Tautomerism of Uracil	17
1.5 DNA Damages	18
1.5.1 Types of Mutations	18
1.5.1.1 Induced Mutations	18
1.5.1.2 Spontaneous Mutations	20
1.6 Deamination of Cytosine	22
1.7 Decomposition Reaction of Formamidine	27
1.8 References	32

Chapter 2 Theoretical Background

2.1 The Schrödinger Equation	56
2.2 Born-Oppenheimer Approximation	58
2.3 The Variational Principle	59
2.4 Slater Determinants	60
2.5 The Hartree-Fock Approximation	61
2.6 Basis Set Expansions	65
2.7 The Post Hartree-Fock Methods	66
2.7.1 Size-Extensivity and Size-Consistency	67

2.7.2 Configuration Interaction	68
2.7.3 Møller-Plesset Perturbation Theory	69
2.8 Density Functional Theory (DFT)	71
2.9 Computational Methodology	75
2.10 G3MP2 Theory	78
2.11 Solvation Models	80
2.11.1 Polarizable Continuum Model (PCM)	80
2.11.2 PCM Calculations	82
2.12 References	85

Chapter 3 *Ab initio* Study of the Decomposition Reaction of Formamidine

3.1 Introduction	91
3.2 Computational Methods	95
3.3 Results and Discussions	96
3.3.1 Thermodynamic Results for the Decomposition of Formamidine	96
3.3.2 Activation Energy, Enthalpy and Free Energies of Activation, for the Decomposition of Formamidine	97
3.3.2.1 The Gas Phase Unimolecular Decomposition of Formamidine	97
3.3.2.2 The Decomposition of Formamidine with	

one Water Molecule	99
3.3.2.3 The Decomposition Reaction of Formamidine	
with two Water Molecules	101
3.3.3 Intermolecular Hydrogen-Transfer Transition States	102
3.3.4 Intrinsic Reaction Coordinate (IRC) analysis	103
3.4 Conclusions	104
3.5 References	106

Chapter 4 Computational Study of the Deamination Reaction of Cytosine with H₂O and OH⁻

4.1 Introduction	126
4.2 Computational Methods	133
4.3 Results and Discussion	134
4.3.1 Deamination of Cytosine with H ₂ O in Pathways A and B	134
4.3.2 Deamination of Cytosine with OH ⁻	137
4.3.2.1 Deamination of Cytosine with OH ⁻ (Pathway C)	138
4.3.2.2 Deamination of Cytosine with OH ⁻ (Pathway D)	140
4.4 Conclusions	142
4.5 References	144

Chapter 5 Mechanisms for the Deamination Reaction of Cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$: A Computational Study

5.1 Introduction	166
5.2 Computational Methods	170
5.3 Results and Discussion	171
5.3.1 Deamination of Cytosine with $\text{H}_2\text{O}/\text{OH}^-$: Pathways A and B	171
5.3.2 Deamination of Cytosine with $\text{H}_2\text{O}/\text{OH}^-$: Pathways C and D	175
5.3.3 Deamination of a Tautomer of Cytosine with $2\text{H}_2\text{O}$ Pathway F ...	178
5.3.4 Deamination of Cytosine with $2\text{H}_2\text{O}/\text{OH}^-$: Pathways G and H	179
5.3.5 Thermodynamic Properties and Relative Stabilities	181
5.4 Conclusions	182
5.5 References	184

Chapter 6 Conclusions and Concluding Remarks203

Appendix A210

Appendix B211

Appendix C212

Appendix D213

Appendix E216

List of Tables

Chapter 3

Table 3.1	Thermodynamic properties for the decomposition reaction of formamidine (in kJ mol^{-1}) at 298.15 K.110
Table 3.2	Activation energies, enthalpies of activation and free energies of activation for the decomposition reaction of formamidine (in kJ mol^{-1}) at 298.15 K.111
Table 3.3	Activation energies, enthalpies of activation and free energies of activation (in kJ mol^{-1}) at 298.15 K for the intermolecular hydrogen-transfers in formamidine/ $2\text{H}_2\text{O}$ (TS4a, TS4b, and TS5).112

Chapter 4

Table 4.1	Thermodynamic properties for the deamination of cytosine with one water molecule (in kJ mol^{-1}) at 298.15 K.160
Table 4.2	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with one water molecule (in kJ mol^{-1}) at 298.15 K (Pathway A).161
Table 4.3	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with one water molecule (in kJ mol^{-1}) at 298.15 K (Pathway B).162

Table 4.4	Thermodynamic properties for the deamination of cytosine with OH ⁻ (in kJ mol ⁻¹) at 298.15 K.163
Table 4.5	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with OH ⁻ (in kJ mol ⁻¹) at 298.15 K (Pathway C).164
Table 4.6	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with OH ⁻ (in kJ mol ⁻¹) at 298K (Pathway D).165

Chapter 5

Table 5.1	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H ₂ O/OH ⁻ (in kJ mol ⁻¹) at 298.15 K (Pathway A).195
Table 5.2	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H ₂ O/OH ⁻ (in kJ mol ⁻¹) at 298.15 K (Pathway B).196
Table 5.3	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H ₂ O/OH ⁻ (in kJ mol ⁻¹) at 298.15 K (Pathway C).197
Table 5.4	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H ₂ O/OH ⁻ (in kJ mol ⁻¹) at 298.15 K (Pathway D).198

Table 5.5	Activation energies, enthalpies of activation and free energies of activation for the deamination of the amino-oxo tautomer of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (in kJ mol^{-1}) at 298.15 K (Pathway F).199
Table 5.6	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with $2\text{H}_2\text{O}/\text{OH}^-$ (in kJ mol^{-1}) at 298.15 K (Pathway G).200
Table 5.7	Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with $2\text{H}_2\text{O}/\text{OH}^-$ (in kJ mol^{-1}) at 298.15 K (Pathway H).201
Table 5.8	Thermodynamic properties and relative stabilities of several species investigated (in kJ mol^{-1}) at 298.15 K.202

List of Figures

Chapter 1

Figure 1.1	The Watson-Crick double helix of DNA where the base pairs are hydrogen bonded.46
Figure 1.2	General structures of pyrimidine and purine.47
Figure 1.3	Structures of the major pyrimidine and purine bases of nucleic acids.48
Figure 1.4	Structures of the minor pyrimidine and purine bases of nucleic acids.49
Figure 1.5	Structures of the ribonucleosides.50
Figure 1.6	Structures of Cytidine 5'-monophosphate.51
Figure 1.7	Watson-Crick base pairing for A · T and G · C.52
Figure 1.8	Six tautomers of cytosine.53
Figure 1.9	Six tautomers of uracil.54
Figure 1.10	Resonance structures of amidinium ions.55

Chapter 3

Figure 3.1	The C_s and C_1 structures of formamidine.118
Figure 3.2	Energy vs the IRC for the gas phase decomposition of formamidine at HF/6-31G(d) with the CalcFC option. The energy of the reactant is taken to be zero.119

Figure 3.3	The transition state structure for the decomposition reaction of formamidine with one water molecule.120
Figure 3.4	Energy vs the IRC for the decomposition of monohydrated formamidine at HF/6-31G(d) with the CalcFC option. The energy of the reactant is taken to be zero.121
Figure 3.5	The transition state structures for dihydrated formamidine.122
Figure 3.6	Energy vs the IRC for the decomposition of dihydrated formamidine, TS3. The energy of the reactant is taken to be zero.123
Figure 3.7	Energy vs the IRC for water/formamidine hydrogen exchange, TS4. The energy of the reactant is taken to be zero.124
Figure 3.8	Energy vs the IRC for water/formamidine hydrogen exchange, TS5. The energy of the reactant is taken to be zero.125

Chapter 4

Figure 4.1	Deamination of cytosine with H ₂ O, deamination of cytosine with OH ⁻ with two possible products, and atom labeling in the amino-oxo tautomer of cytosine.149
Figure 4.2	Comparison of formamidine (<i>E</i> -isomer) and cytosine structure.150

Figure 4.3	Deamination of cytosine with one water molecule to form the Ura-NH ₃ complex (pathway A).151
Figure 4.4	Reaction pathway for the deamination of cytosine with one water molecule at the G3MP2 level of theory (pathway A).152
Figure 4.5	Deamination of cytosine with one water molecule to form the Ura-NH ₃ complex (pathway B).153
Figure 4.6	Reaction pathway for the deamination of cytosine with one water molecule at the G3MP2 level of theory (pathway B).154
Figure 4.7	Deamination of cytosine with OH ⁻ on the <i>si</i> face (see Figure 4.9) to form the Ura anion-NH ₃ complex (pathway C).155
Figure 4.8	Reaction pathway for the deamination of cytosine with OH ⁻ at the G3MP2 level of theory (pathway C).156
Figure 4.9	Two enantiomers of the intermediate structure (I1) corresponding to OH ⁻ attack on the <i>si</i> face and on the <i>re</i> face of the cytosine ring.157
Figure 4.10	Deamination of cytosine with OH ⁻ to form the Ura anion-NH ₃ complex (pathway D).158
Figure 4.11	Reaction pathway for the deamination of cytosine with OH ⁻ at the G3MP2 level of theory (pathway D).159

Chapter 5

Figure 5.1	Deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and the atom labelling in cytosine.186
Figure 5.2	Reaction pathway A with the optimized structures for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory).187
Figure 5.3	Reaction pathway B with the optimized structures for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory).188
Figure 5.4	Comparison of reaction pathways for the deamination of cytosine with OH^- (dotted line) [3] and reaction pathway B for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ at G3MP2 level of theory.189
Figure 5.5	Reaction pathway C with the optimized structures for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory).190
Figure 5.6	Reaction pathway D with the optimized structures for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory). $\text{I4}_\text{D}(1)$ optimized structure at B3LYP/6-31G(d) and $\text{I4}_\text{D}(2)$ at B3LYP/6-31+G(d) and HF/6-31G(d)191

Figure 5.7	Reaction pathway F with the optimized structures for the deamination of the amino-oxo tautomer of cytosine with 2H ₂ O (Relative energies at G3MP2 level of theory).192
Figure 5.8	Reaction pathway G with the optimized structures for the deamination of cytosine with 2H ₂ O/OH ⁻ (Relative energies at G3MP2 level of theory).193
Figure 5.9	Reaction pathway H with the optimized structures for the deamination of cytosine with 2H ₂ O/OH ⁻ (Relative energies at G3MP2 level of theory).194

List of Schemes

Chapter 1

Scheme 1.1	Deamination reaction catalysed by cytosine.41
Scheme 1.2	Deamination reaction of cytosine with sulfite.42
Scheme 1.3	Intramolecular hydrogen-transfer rearrangement of formamidine.43
Scheme 1.4	Water-mediated intramolecular hydrogen-transfer rearrangement of formamidine.44
Scheme 1.5	The unimolecular decomposition reaction of formamidine.45

Chapter 3

Scheme 3.1	The gas-phase unimolecular decomposition reaction of formamidine.113
Scheme 3.2	The decomposition reaction of formamidine in solution.114
Scheme 3.3	The decomposition reaction of formamidine with one water molecule.115
Scheme 3.4	The decomposition reaction of formamidine with one water molecule in solution. 116
Scheme 5.3	The decomposition reaction of formamidine with two water molecules. 117

List of Abbreviations and Symbols

A or (Ade)	Adenine
AID	Activation-induced deaminase
AM1	Austin Model 1
AMP	Adenosine 5'-monophosphate
AO	Atomic orbital
ATP	Adenosine triphosphate
B3LYP	Becke3-Lee-Yang-Parr
5-BU	5-Bromouracil
C or (Cyt)	Cytosine
CASSCF	Complete active space self-consistent field
CBN	Commission on biochemical nomenclature
CC	Coupled cluster theory
CCSD(T)	Coupled cluster single and double substitution operators with perturbative estimate for connected triples
CI	Configuration interaction
CISD	CI with single and double electronic excitations
CMP	Cytidine 5'-monophosphate
COSMO-RS	Conductor-like-screening model for real solvent
CPCM	Conductor PCM calculations
dAMP	Deoxyadenosine 5'-monophosphate
dCMP	Deoxycytidine 5'-monophosphate

DFT	Density Functional Theory
dGMP	Deoxyguanosine 5' – monophosphate
DNA	Deoxyribonucleic acid
dsDNA	Double stranded Deoxyribonucleic acid
dTMP	Deoxythymidine 5' – monophosphate
DZP	Double zeta polarized (basis set)
fc	Frozen core
5-FC	5-Fluorocytosine
5-FU	5-Fluorouracil
G	Guanine
G2	Gaussian-2-theory
G3MP2	Gaussian-3-theory using MP2 to estimate basis set extension effects
GMP	Guanosine 5' – monophosphate
GTF	Gaussian type function
GTO	Gaussian type orbital
HF	Hartree-Fock
I	Intermediate
IRC	Intrinsic reaction coordinate
IUPAC	International union of pure and applied chemistry
KDa	kilodalton
LCAO	Linear combination of atomic orbitals
Mbp	Million base-pairs

MCSCF	Multiconfiguration self-consistent-field
MO	Molecular orbital
MP n	Møller-Plesset perturbation theory of order n
MP2	Second order Møller-Plesset perturbation theory
MP4(SDTQ)	MP4 including single, double, triple and quadruple excitations
MRCI	Multireference CI
NMR	Nuclear magnetic resonance
PCM	Polarized continuum model
PES	Potential energy surface
pKa	Acid dissociation constant
PM3	Parameterized model 3
Ps	Picoseconds
QCI	Quadratic configuration interaction
QCISD	QCI including singles and doubles
QCISD(T)	QCISD with perturbative estimate for connected triples
R	Reactant
RHF	Restricted Hartree-Fock
RNA	Ribonucleic acid
SCF	Self-consistent-field
SCRf	Self-consistent reaction field
STO	Slater type orbital
T	Thymine
tRNA	Transfer Ribonucleic acid

TS	Transition State
U or (Ura)	Uracil
UA0	United atom
UFF	Universal (united) force field
UMP	Uridine 5'-monophosphate
UV	Ultraviolet
YCD	Yeast cytosine deamination
ZPE	Zero-point energy

List of Publications From This Thesis

1. Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A. “*Ab initio* Study of the Decomposition Reaction of Formamidine”, *Canadian Journal of Chemistry* 2005, 83, 2082–2090. **(Chapter 3).**
2. Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A.; Sokalski, W. A. “Computational Study of the Deamination Reaction of Cytosine with H_2O and OH^- ”, *Journal of Physical Chemistry A* 2006, 110 (26), 8227–8234. **(Chapter 4).**
3. Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A.; Sokalski, W. A. “Mechanisms for the Deamination Reaction of Cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$: A Computational Study”, Submitted to the *Journal of Chemical Information and Modeling* on August 29th, 2007. **(Chapter 5).**

CHAPTER 1

GENERAL INTRODUCTION

1.1 Overall objectives and outline

Quantum chemistry has successfully predicted a range of ground state properties, especially regarding organic molecules. Computational chemistry is used to study the structure, properties and reactions of scientifically interesting systems. Such calculations often lead to new experiments and further our understanding of these systems. The study of deoxyribonucleic acid (DNA) has been a high priority since its discovery in the 1950s. Nucleic acids have been prime targets of chemical, medical, physical, and biological investigations. The study of nucleic acid structure has grown into several fields which continue to play a fundamental role in furthering our understanding of gene regulation and expression. DNA is a chemical storage form of genetic information. Ribonucleic acid (RNA) is involved in the transfer of genetic information and in biological catalysis. It has become a central focus of crystallographers, NMR spectroscopists, and theoreticians, and now even has its own

database¹ for structure and study.

DNA can be damaged in several ways, for example, by mutagens. Numerous studies have been carried out on DNA damage and repair. Understanding the details of DNA damage and repair may be used in a disease control. Damage to DNA can be caused by deamination of nucleic acid bases. The hydrolytic deamination of cytosine yields uracil, which is found in RNA only. This results in base pair changes from CG to UG, or from GC to GU (C: cytosine; G: guanine; and U: uracil). If uracil is found in DNA, it is removed by uracil-N-glycosylase. However, the mechanism of the hydrolytic deamination reaction of cytosine in DNA is still unknown. For this reason, we have performed a detailed study of possible mechanisms for the deamination reaction of cytosine with H_2O , OH^- , $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$.

This thesis serves several purposes. First and foremost, it is aimed towards finding a mechanism for the hydrolytic deamination reaction of cytosine which will account for the observed experimental activation energy. Secondly, to investigate the decomposition reaction of formamidine as a contribution to the understanding of the mechanism of proton transfer reactions in hydrogen-bonded systems. Thirdly, it is designed to determine the computationally least expensive level of theory which will give reliable energetic results by performing high level *ab initio* calculations on the formamidine structures. Formamidine is a simple model for hydrogen shift reactions and protonation and deprotonation in the nucleic

¹<http://ndbserver.rutgers.edu/>

acids bases. The intramolecular hydrogen-transfer in formamidine may be considered as a model for the intramolecular hydrogen rearrangement of cytosine. Moreover, formamidine forms part of the cytosine structure. The deamination reaction of formamidine was considered as a model for the study of the deamination of cytosine.

This thesis is in manuscript format which contains six chapters. Each chapter has its own references except the last chapter which summarizes a general conclusion. The thesis is organized as follows. **Chapter 1** presents an introduction and background about the nucleic acid, cytosine deamination, and formamidine decomposition, and is divided into seven sections. In the section titled overall objectives and outline, the main purpose of this thesis along with the outline of the thesis are described. The second section presents an introduction about the nucleic acids which contains four subsections. The first subsection gives a description of the nucleic acids (DNA and RNA). Next, the nucleic acid bases are discussed. In the last two subsections, nucleosides, nucleotides, and base pairing are discussed. In the section titled cytosine, the chemical and physical properties and cytosine tautomerism are discussed as well as a more detailed literature review is presented. Accordingly, the section titled uracil will also be divided into two subsections. The first subsection will focus on the chemical and physical properties and the second will review uracil tautomerism. The fifth section provides some examples of DNA damage with a brief detail about the induced and spontaneous mutations. In the sixth section, the deamination of cytosine is discussed in detail and a detailed description of the literature review is presented. The last section of Chapter 1

gives an introduction and a literature review about the decomposition reaction of formamidine.

Chapter 2 provides a theoretical background which includes twelve sections. The first six sections present an introduction and a brief background on electronic structure theory including the Schrödinger equation, Slater determinants, Hartree-Fock theory, and basis sets expansions. The seventh section gives a definition of electron correlation and presents some examples of post Hartree-Fock theories. The eighth section gives a brief introduction to density functional theory. The computational methods employed for optimization of the geometries and IRC analysis are described in section nine. The G3MP2 calculations are discussed in the section titled G3MP2 theory. In the last section, the solvation methods are described. Some calculations using the G3MP2 level of theory and the solvation methods are included in appendices C and D. **Chapter 3** describes a detailed study of the decomposition reaction of formamidine in the gas phase and in solution and provides a conclusion summarizing the results of this chapter. **Chapter 4** describes a detailed study of the deamination reaction of cytosine with H_2O and OH^- and gives a conclusion summarizing the results of this chapter. **Chapter 5** shows, for the first time, a plausible mechanism for the deamination reaction of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$ and includes a conclusion summarizing the results of this chapter. Finally, a summary and main conclusions about the results of this study are presented in **Chapter 6**.

1.2 The Nucleic Acids (DNA and RNA)

DNA has been the subject of numerous chemical, physical, and biological investigations. The study of nucleic acid is more than 50 years old. Nucleic acids play a fundamental role in the transformation of genetic information. The total genetic information content of each cell, referred to as the genome, resides in the long, coiled macromolecules of DNA. Nucleic acids are of much interest due to their role in the storage, transfer and expression of genetic information. The two fundamental types of nucleic acids are DNA and RNA. DNA is found primarily in the chromosomal form in the cell's nucleus, where it serves as the molecular repository of genetic information. DNA forms chains of genetic material organized into chromosomes. Each chromosome contains a single long molecule of DNA.

The biochemical importance of nucleosides and nucleotides became more apparent in the early 1950s. The era of modern biology was introduced in 1953 when J. O. Watson and F. H. C. Crick determined the structure of DNA [1]. Using X-ray diffraction data and models, they discovered that the DNA molecule is constructed of two strands forming a double helix as shown in Figure 1.1. DNA is a very narrow thread with a diameter of just $1-2 \times 10^{-9}$ m [2]. DNA molecules are the largest of the macromolecules found in a cell. For example, the chromosome in *E. Coli* has 4 Mbp (million base-pairs) and a length of 1.5 mm [3]. The mass of this molecule is 2.6×10^6 kDa (kilodalton). The genomic DNA of a single human cell has 3900 Mbp and is 990 mm long [3]. The nucleic acids are heteropolymers, made up of a linear array of monomeric units called nucleotides. Different nucleic acids can form between

approximately 80 nucleotides- as in tRNA- and over 10^8 nucleotide pairs in a single eukaryotic chromosome. Each monomer unit consists of three parts: an organic base; a carbohydrate (ribose or deoxyribose); and a phosphate. The double helix consists of two complementary base pairs twisted to form a circular staircase-type structure. The two strands run in opposite directions, anti-parallel (see Figure 1.1). The Watson-Crick model for the double helix has several conformational structures (such as: A-DNA, B-DNA, and Z-DNA) that have been experimentally observed. These double helices have different number of phosphates per turn of helix for each strand. The A-DNA and B-DNA are right-handed with 11 and 10 phosphates per helical turn, respectively. The left-handed Z-DNA has 12 phosphates per helical turn [2,3]. These slight differences are very significant in biology. The favored form at low humidity (and high salt) is the highly crystalline A-DNA. The dominant structure is B-DNA at high humidity and low salt [3].

1. 2. 1 The Nucleic Acid Bases

All of the bases in DNA and RNA are derivatives of the two heterocyclic compounds purine and pyrimidine. Purine and pyrimidine, as shown in Figure 1.2, are nitrogen containing ring compounds of biological significance. The major six-membered monocyclic pyrimidines in DNA are cytosine (Cyt or C), uracil (Ura or U), and thymine (Thy or T). The major bicyclic purines are adenine (Ade or A) and guanine (Gua or G), and are found in both DNA and RNA [3-6]. All are components of DNA and adenine is a component of ATP (Adenosine triphosphate). Figures 1.3 and 1.4 show some major and minor heterocyclic bases in DNA and

RNA along with the official numbering systems. The pyrimidines in RNA are cytosine and uracil. Uracil usually replaces thymine in RNA and differs from thymine by loss of a methyl group on its ring. Throughout the thesis, we will use the abbreviation and symbols for nucleic acid, as recommended from the IUPAC-IUB Commission on Biochemical Nomenclature (CBN)².

1. 2. 2 Nucleosides and Nucleotides

The combination of the purine or pyrimidine base with a pentose sugar is called a nucleoside. Two numbering systems are necessary to identify carbon and nitrogen atoms. Primed numbers (1', 2', 3',...,etc) are used for the pentose ring. The binding is a glycosidic bond from the 1' carbon of the sugar to N₁ of the pyrimidine or N₉ atom of the purine as shown in Figure 1.5. The combination of nucleic base, sugar, and phosphoryl group ($-PO_3^{2-}$) is called a nucleotide, as shown in Figure 1.6. Nucleotides are the phosphate esters of nucleosides and these are components of both RNA and DNA. The linear structure of DNA occurs when deoxyribose, the sugar in DNA, binds with a phosphate at both its 3' and 5' carbons. RNA is made up of ribonucleotides while the monomers of DNA are 2'-deoxyribonucleotides. The most common site for the ester linkage is the hydroxyl group on carbon 5' of the pentose (see Figure 1.6). DNA is composed of the monomers deoxyadenosine 5'-monophosphate (dAMP), deoxyguanosine 5'-monophosphate (dGMP),

²<http://www.chem.qmul.ac.uk/iupac/misc/naabb.html>

deoxycytidine 5'-monophosphate (dCMP), and deoxythymidine 5'-monophosphate (dTMP). Ribonucleic acid is involved in the transfer of genetic information and in biological catalysis. RNA is a heteropolymer of adenosine 5'-monophosphate (AMP), guanosine 5'-monophosphate (GMP), cytidine 5'-monophosphate (CMP), and uridine 5'-monophosphate (UMP). DNA in the human chromosome contains millions of nucleotides. The phosphate group is acidic, so nucleotides are acids. Nucleotides are negatively charged because of charges carried on the phosphate. In DNA and RNA, the nucleotides are linked by 5'-3' phosphodiester linkages. A DNA sequence is always read from 5' to 3' [4-6]. Nucleic acids have acidic and basic regions. The phosphate groups of the backbone have negative charges because of proton dissociation at physiological pH. The ionic character of the many negative charges and a polar sugar components establish a hydrophilic region that interacts favorably with water. The acid-base behavior of a nucleotide determines its charge, its tautomeric structure, and thus its ability to give and accept hydrogen bonds. The genetic information present in nucleic acids is coded by the sequence of nucleotides [6].

1. 2. 3 Hydrogen Bonding: Base Pairing

Multiple hydrogen bonding plays a significant role in a number of biologically important phenomena; e.g., the base-pair interactions in DNA which are very important for the transfer and expression of genetic information. Löwdin [7, 8] suggested that hydrogen-bonded protons in base pairs may be involved in the formation of unusual imino-enol tautomers and may play a key role in spontaneous mutations. Since then, a great deal of

theoretical work [7–15] has been done to study the hydrogen-bonding of the base pairs in DNA.

The nitrogen bases have several functional groups that form hydrogen bonds. (N–H) groups and amino groups ($-\text{NH}_2$) are potential hydrogen bond donors. Carbonyl groups (sp^2 carbon) and ring nitrogen atoms (sp^2 nitrogen) ($=\text{N}-$) are potential hydrogen bond acceptors. Specific hydrogen bonding occurs between the bases of one strand and the bases of another strand. This base-pairing pattern became known as Watson-Crick pairing. Purines form hydrogen bonds to pyrimidines: for example, adenine bonds only to thymine via two hydrogen bonds; and cytosine bonds to guanine via three hydrogen bonds (i.e., $\text{C} \equiv \text{G}$, $\text{T} = \text{A}$, $\text{U} = \text{A}$). Figure 1.7 shows the Watson-Crick base-pairing for $\text{C} \cdot \text{G}$ and $\text{T} \cdot \text{A}$ [3]. The hydrogen bond in ($\text{C} \equiv \text{G}$) is stronger than ($\text{T}/\text{U} = \text{A}$) because they stack better and have one more hydrogen bond. For isolated pairs, the $\text{C} \cdot \text{G}$ pair has stronger binding energy compared to the $\text{T} \cdot \text{A}$ pair ($\text{C} \cdot \text{G} = -87.8 \text{ kJ mol}^{-1}$ versus $\text{T} \cdot \text{A} = -54.3 \text{ kJ mol}^{-1}$) when measured in a vacuum [5]. The double helix is stabilized by two types of forces: (1) hydrogen bonds between pairs of complementary bases on opposite strands; (2) Van der Waals and hydrophobic interactions between stacked bases.

Base stacking interactions affect the three-dimensional structure of DNA. The hydrogen-bonded interactions are believed to be governed mainly by the electrostatic and dispersion terms. The stacked complexes of DNA bases are stabilized mostly by the dispersion

energy [16]. Several theoretical studies have been done on base stacking, including the cytosine dimer [12, 16–20]. Sponer et al. [12] studied the nature of nucleic acid-base stacking by *ab initio* and empirical potential calculations. The interaction energies were calculated for 240 geometries of 10 stacked nucleic acid-base pairs. They found that the most stable stacked pair is the G...G dimer ($-47.2 \text{ kJ mol}^{-1}$), and the least stable is the U...U dimer ($-27.2 \text{ kJ mol}^{-1}$).

1.3 Cytosine

1.3.1 Chemical and Physical Properties

Cytosine has two possible chemical names, 4-amino-2(1H)-pyrimidinone or 4-aminopyrimidine-2(1H)-one. It is found in both DNA and RNA. The nucleoside of cytosine is called cytidine (Cyd) in RNA and deoxycytidine in DNA. The nucleotide of cytosine is cytidine 5'-monophosphate (5'-CMP) in RNA and deoxycytidine 5'-monophosphate (5'-dCMP) in DNA. The pKa values for cytosine in the nucleoside, 3'-nucleotide, and 5'-nucleotide are 4.17, 4.43, and 4.56, respectively [3, 6]. The protonation of cytosine most likely occurs on one of the ring nitrogens rather than on the exocyclic amino group. Protonation of cytosine occurs most favourably at the N₃ site [21, 22], where the positive charge can delocalize throughout the N₁, N₃, and exocyclic nitrogens. This makes C₅ of cytidine less reactive to the electrophiles that may be present in acidic medium, as compared with the C₅ in uridine. Nucleic acids contain small quantities of methylated bases such as: 2-methyl adenine and 1-methyl guanine. Cytosine can be methylated and form

5-methylcytosine (m^5C) (by an enzyme called DNA methyl transferase), 1-methylcytosine, and 5-hydroxy methyl cytosine [2, 23, 24] (For more detail about DNA methylation see references 2 and 24). Cytosine can be converted to uracil by spontaneous mutation (deamination). If uracil is found in DNA, it poses a very serious problem; the cell, however, has a specific enzyme called uracil-N-glycosylase to remove it. Uracil glycosylase removes the uracil bases from DNA wherever they are accessible, and then a repair polymerase pairs the opposite guanine base with cytosine in order to fix the damage. Of the nucleic bases, cytosine most readily undergoes nucleophilic attack, especially by nucleophiles such as hydrazine, hydroxylamine,...etc.

Fülscher et al. [25], studied the electronic spectrum of cytosine using CASSCF (Complete active space self-consistent field) and MRCI (Multireference Configuration interaction) calculations. They found that the computed vertical $\pi \rightarrow \pi^*$ transition energies for the four lowest singlet excited valence states are 4.4, 5.4, 6.2, and 6.7 eV, which were in good agreement with experimental values. Florián et al. [26], presented an experimental and theoretical study of the vibrational spectra of cytosine and protonated cytosine. They studied IR and Raman spectra and tautomeric stabilities of protonated cytosine by the *ab initio* (HF, MP2, MP4) and B3LYP levels of theory with the 6-31G(d) basis set. According to their calculations, O_2 -protonated (enol) cytosine prevails slightly in the gas phase, whereas the N_3 -protonated cytosine ($12.5 \pm 4 \text{ kJ mol}^{-1}$) is more stable than the enol tautomer in an aqueous solution [26]. Cytosine and many of its derivatives have been studied theoretically with $\cdot OH$

[27], with halogenated derivatives including hyperfine coupling constants [28], with $\cdot\text{H}$ [21, 22], and with hydrogen peroxide [29]. Recently, Zhang et al. [22] carried out DFT calculations of three possible reactions of $\cdot\text{H}$ with cytosine. They found that the addition of H atom to the N_3 site of cytosine is energetically more favourable than the C_5 or C_6 site. Moreover, the addition of the H atom to the C_5 site is more favourable than the addition to the C_6 site with 8.4 kJ mol^{-1} difference. More recently, Dong et al. [30] studied the interaction between cytosine and BX_3 ($\text{X}=\text{F}, \text{Cl}$) systems. In particular, they studied eight energy minimum conformers, four conformers for cytosine- BF_3 and four conformers for cytosine- BCl_3 using HF and B3LYP levels of theory with 6-311+G(d,p) basis sets. They found that the cytosine- BF_3 complex is more stable than the cytosine- BCl_3 complex [30].

Solvation plays an important role in the tautomeric equilibrium between the “normal” and “rare” forms of nucleic acid bases. The solvent can control the proton transfer reaction in a system with solute/solvent interactions. A large amount of work has been performed on the interaction between cytosine and its tautomers with water molecules [31–36]. Hunter et al. [35] studied the hydrogen bonding properties of cytosine with HF, H_2O and NH_3 . They studied the interactions of these molecules at four main binding sites in neutral and N_1 anionic cytosine. For neutral cytosine, the cytosine- H_2O complex at the O_2-N_3 site is less stable than those involving bidentate hydrogen bonds. They investigated the cytosine-water complex interaction further [36] and they pointed out that by using B3LYP with triple-zeta basis sets, the O_2-N_3 complex can be fully optimized to a ring-opened structure with water coordinated

to the O₂ atom.

It is a well-known fact that the structures and properties of DNA depend on metal ions [37]. Metalation can play an important role in the formation of the canonical and non canonical structures of DNA [38] and the ability of the nucleobase to be protonated or deprotonated. DNA base pairs can interact with metal ions at different sites: phosphate groups, the sugar moiety as well as the DNA moiety. The metal cations can either interact with the nucleic acid bases at the N₇ and the O₆ positions of guanine or N₇ position of adenine. The nucleobase-metal interactions at N₃ of cytosine and N₁ of adenine are not accessible in double-stranded DNA under Watson–Crick base pairing [38]. Several papers have discussed the interaction of nucleic acid bases with metal cations including the metalated cytosine [37–43]. Lee and co-workers, in 1993, proposed that by adding Zn²⁺ to a solution of DNA at high pH, the M-DNA structure (where M stands for metal) will be produced [43]. More recently, the structure, stability and electronic properties of the G-Zn-C base pair in M-DNA has been studied [43].

More recently, the ultrafast radiationless decay mechanism of photo excited cytosine (keto-tautomers) has been studied theoretically using the MRCI *ab initio* method [44]. It has been reported that the excited-state life time of cytosine ranges from 0.72 to 3.2 ps. [44]. It has been found that the damage to DNA fragments from very low energy (<1 eV) electron can occur. DNA bases can attach electrons having energy below 1 eV to their π^* orbitals and

undergo phosphate-sugar bond cleavage (o-c) at rate of 10^6 s^{-1} at $T = 298 \text{ K}$ or the loss of an H atom directly from the base's ring at rates of 10^5 s^{-1} .

1. 3. 2 Tautomerism of Cytosine

Tautomerism of nucleic acid bases (e.g., cytosine, thymine, and uracil) is of great importance to the structure and functioning of nucleic acids and has generated considerable interest due to the proposal that the occurrence of certain tautomers can lead to spontaneous mutation in the genetic code in DNA or RNA helices [7]. A great deal of work has been carried out on the tautomerism of cytosine using both experimental [26, 45–47] and theoretical [26, 31–33, 48–57] methods. Some of the literature has been reviewed in the introduction of Chapter 4. Numerous computational studies have discussed the tautomerism of a cytosine molecule in both the gas phase [26, 45, 49–53, 56, 57] and the aqueous phase [34, 52–55]. Cytosine may exist in different tautomeric forms which differ from each other mainly in the position of one of the hydrogens which may be bound to the exocyclic amino nitrogen or oxygen atom or to ring nitrogen atoms. Figure 1.8 shows the six cytosine tautomers, the (a-o) amino-oxo (C1), (i-o) imino-oxo (C2), (a-h) (*cis*) amino-hydroxy (C3), (a-o) amino-oxo (C4), (i-h) (*cis*) imino-hydroxy (C5), and (i-h) (*trans*) imino-hydroxy (C6). Experimentally, it has been found that cytosine exists as a mixture of C1 (a-o)- the canonical form- and the C2 (a-h) tautomers [45, 58].

The (a-o) form of cytosine is the predominant tautomer in both the solid and in the

solution [51, 55]. The first investigation of the cytosine tautomers by *ab initio* methods [59] predicted C3 to be the most stable tautomer which is in disagreement with experiment. Scanlan et al. [92] studied the tautomerism of uracil, thymine, 5-fluorouracil and cytosine. They found that C1 is the dominant tautomer in aqueous solution. They also found that the order of stability for the tautomers in aqueous solution is $C1 > C4 > C2$ while C3, C5 and C6 are not detected. In addition, they predicted the order of stability in gas phase to be $C1 > C2 > C3 > C4$. In the case of C3 tautomer, there are two conformers that exist, in which the O-H bond being *cis* and *trans* with respect to the N_1-C_2 bond. The *cis* conformation is more stable than *trans* by 3.4 kJ mol^{-1} [51, 57]. All calculations carried out on the tautomerism of cytosine assumed all tautomers to be planar except the work done by Leś [57]. Leś et al. [57] studied the C1, C2 and C3 forms of cytosine by the coupled cluster (CC) method with DZP (Double zeta polarized) basis sets. They reported that the (a-o) and (i-o) forms are less stable than the (a-h) tautomer by 4.2 and 6.4 kJ mol^{-1} , respectively. Estrin [50] found the C1 tautomer to be the most stable tautomer by using the DFT calculations. Recently, Sambrano et al. [55] studied cytosine tautomerism in both the gas phase and in solution (using continuum models) by B3LYP/6-31++G(d,p) and MP2/6-31++G(d,p) levels of theory. In the gas phase, they found that the order of stability was $C3 > C1 > C2 > C4 > C5 > C6$ with an energy difference between C3 and C1 of 1.53 kJ mol^{-1} at MP2. At B3LYP, the order of stability was: $C1 > C3 > C2 > C4 > C5 > C6$ with an energy difference between C1 and C3 of 1.7 kJ mol^{-1} . In aqueous media, both levels predicted the C1 tautomer to be the most abundant form. They also found that the amino group becomes ring-coplanar in aqueous medium at the B3LYP

level of theory.

1. 4 Uracil

1. 4. 1 Chemical and Physical Properties

Uracil has three possible chemical names: 2-oxy-4-oxypyrimidine, pyrimidine-2,4(1H,3H)-dione, and 2,4-dihydroxypyrimidine. Uracil forms hydrogen bonds with adenine and is replaced by thymine in DNA, where thymine is produced by methylation of uracil. The pKa values for uracil in the major nucleosides, 3'- nucleotide and 5'- nucleotide are 9.38, 9.96 and 10.06, respectively [3, 6]. The combination of uracil and ribose sugar is called ribonucleoside, uridine (Urd). When uridine combines with a phosphate, uridine 5'-monophosphate is produced. Uridine has the least basic character among the naturally occurring nucleosides. If the nucleophile has a basic character, deprotonation of the N₃ of uridine makes nucleophilic attack less favourable than that observed for cytidine [6].

Other examples of the uracil derivatives are 5-bromouracil (5-BU) and 5-fluorouracil (5-FU) which are two biologically active pyrimidines. 5-FU is an example of the deliberate use of C-F bonds in drug design and 5-FU, an analog of thymine, is an inhibitor of DNA synthesis and has been widely used in cancer treatment (anticancer drug).

1. 4. 2 Tautomerism of Uracil

Uracil exists in six possible tautomeric forms (U1, U2,..., U6), as shown in Figure 1.9. Several theoretical studies on uracil have been reported [34, 50– 52, 60] due to the fact that the formation of rare tautomers can lead to spontaneous mutations . Scanlan [51] found that U3 tautomer is less stable than U1 tautomer by 81.8 kJ mol^{-1} in the gas phase using 3-21G basis set. Leszczynski [60] studied the tautomerism of uracil at MP2/6-31G(d,p)//HF/6-31G(d,p) and single point calculations at MP4(SDTQ) and CISD using HF/6-31G(d,p) optimized structures. U1 is predicted to be the predominant tautomer. DFT calculations show that the U1 (dioxo or amino-oxo tautomer) is the most stable tautomer. The order of stability is $U1 > U2 > U3 > U4 > U5 > U6$. Semiempirical AM1 and PM3 calculations [52] showed that U1 tautomer or (a-o) form is the most stable tautomer in both the gas phase and aqueous solution.

More recently, the effects of hydrogen bonding on the acidity of uracil have been studied [61,62]. Laudo et al. [61] studied the effect of hydrogen bonding on the acidity of uracil and uracil anion complexes with water, ammonia, and hydrogen fluoride at various uracil sites ($O_2(N_3)$, $O_4(N_3)$ and $O_4(C_5)$). They found that the binding strength of (H_2O , NH_3 , and HF) molecules to neutral uracil is largest at the $O_4(N_3)$ position and at the $O_2(N_3)$ position in the uracil anion.

1.5 DNA Damage

DNA contains all the information necessary to make proteins that perform normal biological functions (cellular functions). Any errors that occur during the replication of DNA must be repaired in order to prevent them from passing to future generations. Changes in the base sequence of DNA are called mutations. Mutations may occur by spontaneous processes or by induced processes, depending on the cause of the damage [63, 64]. Mutated genes produce proteins that are involved in some abnormal activity and show loss of cellular functions. Nucleic acid bases can be altered either by environmental or chemical means. Once altered, these changes may then be propagated by further DNA replication. Cell DNA repair systems prevent mutation from being replicated and transferred to daughter cells. More recently, Berti et al. [65] wrote a review article which represents a detailed understanding of excision repair enzymes.

1.5.1 Types of Mutations

1.5.1.1 Induced Mutations

Organisms are often exposed to mutagens that produce mutations. Mutagens are agents that are capable of causing changes in the DNA base sequence (mutations). Some common examples of these mutations are:

(1) Exposure of cells to ionizing radiation (gamma rays or X-rays) causes double strand

breaks in the DNA and it can also generate reactive oxygen (radicals).

(2) Exposure of cells to UV light can cause adjacent pyrimidines on the same strand to dimerize, in which case thymine dimers or thymine-cytosine dimers can result. These dimers cause DNA polymerase to stop functioning. This can occur by forming a cyclobutane ring which links carbons 5 and 6 of each pyrimidine ring to one another. For a review on this topic, see Ref. [19] and references therein.

(3) Reactive oxygen (H_2O_2 , OH^-) which is generated by ionizing radiation and chemical agents gives rise to transversions. Transversion mutations change the chemical nature of the base. Thus, a pyrimidine base is replaced by a purine base and vice versa ($\text{A} \rightarrow \text{C}$ or $\text{A} \rightarrow \text{T}$), see references [27, 29].

(4) Chemical mutagens are chemicals which cause genetic mutations. For example, nitrous acid (HNO_2) deaminates cytosine residues with formation of uracil. Alkylating agents (nitrosamines, dimethyl sulfate, nitrosoguanidines) are mutagenic because they add methyl or ethyl groups to base residues. Intercalating agents, ethidium bromide and acridine, can stack with the bases and shift the code. Nitrous acid oxidizes the amino groups on nucleotide bases converting guanine to xanthine, adenine to hypoxanthine, cytosine to uracil, and 5-methylcytosine to thymine. These changes cause mispairing and when not corrected, lead to transition mutations.

1. 5. 1. 2 Spontaneous Mutations

Spontaneous mutations occur during normal genetic and metabolic functions in the cell. They can arise as a result of chemical reaction to individual bases in DNA or are associated with DNA replication. Some examples are:

(1) Tautomerization:

It is a rearrangement of protons in the DNA base. These mistakes must be repaired by the mismatch repair system. For example, the bases uracil and cytosine exist in several possible keto-enol tautomeric forms. Tautomerism in cytosine and uracil have been extensively studied theoretically [26, 31–33, 48–57, 60].

(2) Deamination of bases:

A deamination reaction, in general, refers to the conversion of an amino group to a carbonyl functional group (the loss of an amino group of a molecule). Deamination of cytosine, adenine, and guanine rarely occur. All the bases, except T and U have amino groups, but only the deamination of cytosine gives a base found in DNA/RNA. Spontaneous deamination of guanine yields xanthine which will base pair with cytosine but with two hydrogen bonds instead of three. Spontaneous deamination of adenine gives rise to hypoxanthine which can now base pair with cytosine. Spontaneous deamination of cytosine will be discussed in a separate section (1.6).

(3) Alkylation:

Alkylation of DNA can result in mutations in several ways. In alkylation, methyl or ethyl groups (ethyl methane sulfonate, ethyl nitrosourea) are transferred to reactive sites on the bases and to phosphate in the DNA backbone [24].

(4) Base analogs:

Base analogs such as 5-bromouracil are analogous to thymine which can base pair with guanine.

1.6 Deamination of Cytosine

The hydrolytic deamination reaction of cytosine leads to the formation of uracil as shown in Scheme 1.1. The deamination of cytosine, its nucleosides, and its nucleotides has been studied extensively by experimental [66–78, 80–87] and theoretical [79–88] methods. The rate of the deamination of cytosine had been calculated from chemical measurement [67, 71, 72] of the amount of uracil produced, and using a sensitive genetic assay [73]. It has been reported that deamination of cytosine is an extremely rare event under normal physiological conditions ($\text{pH} = 7.4$). Several studies have suggested [66, 67] that cytosine is deaminated in neutral and weakly acidic buffers. Shapiro and Klein [67] stated that cytosine and cytidine are slowly deaminated to uracil and uridine, respectively, by hot carboxylate or pyridine buffers at $\text{pH} < 6$ and $95\text{ }^{\circ}\text{C}$. They also reported that adenosine and guanosine are not deaminated under these conditions. They proposed that the cytosine deamination might have mutagenic effects which are repaired by DNA uracil-glycosylase. The rate of deamination can be increased in the presence of various reagents such as bisulfite, NO , and HNO_2 . The deamination of cytosine derivatives by sodium bisulfite was studied and first proposed by Shapiro [68, 70]. Shapiro [68] reported an important feature about the cytosine and uracil chemistry which is the susceptibility of their ring to nucleophilic attack at the C_5 and C_6 position ($\text{C}_5 = \text{C}_6$) [69]. Scheme 1.2 shows deamination of cytosine catalysed by sulfite [68, 70]. The bisulfite-induced deamination has been studied in acid and base-catalysed mediums [68–70, 72, 74].

The deamination of 5-methylcytosine (m^5C) yields thymine, which is found in DNA [66]. Replication of the deamination products for cytosine and m^5C produce a C·G → T·A transition mutation. The bisulfite method has proven to be very useful to distinguish normal cytosine from m^5C content in clinical DNA samples, in order to determine whether or not a person has a particular disease [2]. Cytosine reacts well with $NaHSO_3$ at N_6 position and then it converts to uracil at high pH. However, m^5C will not react with sodium bisulfite, due to steric hindrance between sulfite at C_6 position and methyl group at C_5 position [2, 66–70]. If the cytosine is m^5C , deamination will yield thymine which is naturally found in DNA. Therefore, methylated C_5 is a hotspots for spontaneous hydrolysis of the N-glycosyl linkage, producing a basic site in the DNA. There is another repair enzyme called thymine-glycosylase that recognize the base-pair mismatch of TG on one strand with CG on the other strand, then cuts out the thymine and replaces it by cytosine with the help of a repair polymerase [2].

Frederico et al. [73], measured the cytosine deamination rate constants in single stranded DNA at temperatures ranging from 30 to 90 °C by a sensitive genetic assay. Under physiological conditions (37 °C and pH 7.4), their calculated cytosine deamination rate for single- and double-stranded DNA is 1×10^{-10} and about $7 \times 10^{-13} \text{ s}^{-1}$ respectively, with an activation energy of $117 \pm 4 \text{ kJ mol}^{-1}$. The measured half-life for cytosine in DNA at 37 °C is 200 years, while in double-stranded DNA (dsDNA) it is on the order of 30 000 years. They reported that cytosine deamination in the double helix is ~140-fold slower than in single-stranded DNA. Their reported activation energy agrees well with the previous value of 121

kJ mol^{-1} by Lindahl and Nyberg [71] who investigated the rate of deamination of cytosine residues in single- and double-stranded *Escherichia coli* DNA as a function of temperature, pH, and buffer composition. They found that the rate of cytosine deamination is $2 \times 10^{-7} \text{ s}^{-1}$ at 95°C and pH 7.4. Other studies reported that hydrolytic deamination was pH and temperature dependent [67, 71].

For yeast cytosine deaminase (YCD), a zinc metalloenzyme, has been studied experimentally [75–78] as well as computationally [79]. YCD catalysed the hydrolytic deamination of cytosine to form uracil. Yao et al. [75], studied the deamination of cytosine and prodrug 5-fluorocytosine (5FC) to form the anticancer drug 5-fluorouracil (5FU). They found from kinetic and NMR results that the release of 5FU is the rate-determining step in the activation of the prodrug 5FC by YCD and may involve multiple steps [75]. There also exist several possible routes for the deamination of the DNA bases including diazotization [80, 81] and nitrosative mechanisms [82, 83]. Glaser et al. [80, 81], proposed that the deamination of the DNA bases cytosine, adenine and guanine can be achieved by way of diazotization and the diazonium ions of the DNA bases. Snider et al. [78], studied the ^{15}N kinetic isotope effects on uncatalyzed and enzymatic deamination of cytidine. They state that the uncatalyzed and the enzymatic reaction proceed by a similar mechanism which involves the formation of a tetrahedral intermediate by stepwise addition of OH and elimination of NH_3 .

Recently, Rayat [82] studied nitrosative cytosine deamination involving pyrimidine

ring-opened intermediates by experimental and theoretical methods. They proposed a possible mechanism for uracil formation which involves ring-opening, water addition and reclosure by condensation steps. Cytosine catalysis of the nitrosative guanine deamination has been studied using HF and B3LYP with the 6-31G(d) basis set [83]. More recently Samaranayake et al. [84] (see references therein) reviewed the role in cytosine deamination by activation-induced deaminase (AID), which is essential for an effective immune response and DNA repair.

Sponer et al. [88], studied the metal-mediated deamination of 1-methylcytosine and 1,5-dimethylcytosine with a cationic complex of Pt^{II} both experimentally and using DFT calculations. As a part of their investigation, they also studied the deamination of cytosine with OH^- using the PCM solvent model. This paper represents the first computational study of the hydrolytic deamination of cytosine. However, their results have shown that the activation energy for this system is $213.4 \text{ kJ mol}^{-1}$ at B3LYP/6-31G(d) using PCM model. The starting point for the mechanism in their paper and our work is the same. However, their mechanism is different from our work in a number of ways. For more detail, see Chapter 4. The potential energy surface for this system shows that the intermediate and the second transition state have almost the same energy relative to the reactant (Cyt anion/ H_2O) complex. They reported two transition states which IRC analysis reveals are not connected in the reaction mechanism. This is because one belongs to one pathway for OH^- attack on the right side of cytosine (*re* face) and to another pathway for OH^- attack on the left side of cytosine

(*si* face). Chapter 4 shows two pathways for the deamination of cytosine with OH^- in which OH^- attacks both faces of cytosine. Both pathways involve two key steps connected with several conformers.

More recently, Matsubara et al. [79], performed a computational study on the catalyst cytidine deaminase and the contribution of a second water molecule. As a part of their investigation, they studied the uncatalyzed hydrolytic deamination of cytosine with H_2O . They reported that the catalytic action of cytidine deaminase is effectively enhanced by the participation of the second water molecule. Their activation energy for the deamination with H_2O , which involves a two step mechanism, was $237.4 \text{ kJ mol}^{-1}$. However, the mechanism reported in their paper differs in a number of ways from our work reported in Chapter 4. In our work, the two key steps are similar to their work, however, they did not pay attention to the conformation changes and the possibility of formation of a cytosine tautomer.

1.7 Decomposition Reaction of Formamidine

Hydrogen-transfer reactions are among the most important fundamental reactions in chemistry and are an important problem in biochemical systems. There have been numerous experimental [89–92] and theoretical studies [93–98] on hydrogen-transfer reactions. Results from studies of single-proton transfers or double-proton transfers in hydrogen-bonded systems can be used as a simple model of multi proton transfer. Examples of multi hydrogen transfer (or multi proton transfer) are proton transfer between DNA base pairs and prototropic tautomerism, in which more than one proton is transferred during the reaction. Single- and double-proton transfers have also been studied [89–91,99] since they are important in prototropic tautomerism in solution and in DNA base pairs.

Amidine compounds play an important role in many medical and biochemical processes. Some of them have been reported to exhibit antibiotic, antiviral, antifungal, antibacterial, and anaesthetic properties [100]. Formation of amidine occurs during the biosynthesis of the purine ring (see page 256 in reference [100] for detail). Amidine compounds contain a C = N double bond with an amide-like C–N single bond indicated by the resonance form as shown in Figure 1.10. In addition, they contain two nitrogen basic centers of different hybridization (sp^2 and sp^3). Amidines are strong bases which undergo protonation on the imino (sp^2 nitrogen) yielding a stable amidinium ion due to the resonance stabilization as shown in Figure 1.10.

The amidines may be classified into several general types depending on the substituents number and their distribution on the nitrogen atoms. For example, formamidine (methanimidamine: $\text{H}_2\text{N}-\text{CH}=\text{NH}$) is the simplest unsubstituted type. All these types of amidines show *cis-trans* isomerism with respect to the carbon-nitrogen double bond ($\text{C}=\text{N}$) as well as rotational isomerism around the $\text{C}-\text{N}$ single bond and tautomerization which may occur in monosubstituted and disubstituted amidines (with different substituents on the nitrogen atoms). An early review about the chemistry of amidine was done by Shriner et al. [101]. Unsubstituted amidines possess antiviral properties and are thus of considerable chemotherapeutic interest.

Formamidine is a simple amidine which has been investigated extensively by experimental and theoretical methods because it hydrogen bonds with itself and with water. This relatively small compound is a suitable system for modelling the bases of nucleic acids (e.g., adenine, cytosine). Moreover, the reaction of amidines are of interest because amidine compounds, like formamidine, form part of the molecules of biological interest, e.g, the DNA bases cytosine, guanine, and adenine . Figure 4.2 in Chapter 4 shows the comparison between formamidine and cytosine, in which the *E* isomer of formamidine forms part of the active region of the cytosine structure, where the deamination reaction can take place [102, 103].

Tautomerization by intramolecular and intermolecular 1,3-sigmatropic hydrogen rearrangement (1-3 proton transfer) for formamidine and its complexes have been a prime

target for extensive theoretical investigations [13–15, 104–128]. The introduction of Chapter 3 will review some of these works. Scheme 1.3 shows the intramolecular hydrogen rearrangement of formamidine, which is the simplest example of amino-imine tautomerism. The intramolecular hydrogen transfer in formamidine (1,3-sigmatropic rearrangement) was first studied theoretically by Yamashita et al. [125]. Scheme 1.4 shows the intermolecular hydrogen rearrangement of formamidine with one water molecule (formamidine-water system) [104,105]. The tautomerization of formamidine (intramolecular and intermolecular rearrangements) may be considered a basic model for proton transfers in nucleic acid bases and used to investigate the water-mediated 1-3 proton transfer in these systems.

Since molecular conformation is very important in biochemical systems, several theoretical studies of the *E* (trans) and *Z* (cis) configurations (isomers) of formamidine have been performed [122, 127–130]. Experimental results of the formamidine derivatives suggest the existence of two isomers [131]. Tortajada et al. [122] investigated the relative stabilities of the *E* (trans) and *Z* (cis) isomers of formamidine and the complexes of formamidine with several monocations (H^+ , Li^+ , Na^+ , Mg^+ , and Al^+). They found the *E*-isomer is 7.5 kJ mol^{-1} more stable than the *Z*-isomer at both G2 and G2MP2 levels of theory. Kaushik et al. [127] found that the *E*-isomer is 5.9 kJ mol^{-1} more stable than the *Z*-isomer at B3LYP/6-31G(d) level of theory. The activation energy for the *E*-*Z* isomerization is $111.7 \text{ kJ mol}^{-1}$, which is in good agreement with the G2 result ($104.6 \text{ kJ mol}^{-1}$) [122]. Experimental results of the formamidine derivatives suggest the existence of two isomers. Zhang et al. [119] have studied

the tautomerization of formamidine in the gas phase, as well as in mono-, di-, and trihydrated formamidine systems. They found that adding one, two, or three water molecules to form cyclic hydrogen bonded clusters stabilizes the transition state and reduces this barrier to 91.5, 83.6, 99.1 kJ mol⁻¹, respectively at CCSD(D)/6-31G(d,p)//MP2/6-31G(d,p). More recently, Flinn et al. [102], our research group, studied the deamination of the more stable *E* conformer of formamidine with OH⁻, H₂O, and H₃O⁺ to form formamide. They found that deamination with OH⁻ is the most likely mechanism with a much lower activation-energy barrier (58.2 kJ mol⁻¹) at the G2 level of theory. Recently, several researchers have studied the formamidine-water system [105,119,124,129], where they found that when the proton transfer is mediated by one water molecule (water-mediated double proton transfer), the barrier is reduced by about 84 kJ mol⁻¹.

Amidine is also known to undergo a unimolecular decomposition reaction to yield hydrogen cyanide and ammonia. Scheme 1.5 shows the decomposition reaction of formamidine. The decomposition reaction of formamidine involves a 1-3 proton shift of the hydrogen bonded to the imino nitrogen (sp² nitrogen) to the amino nitrogen (sp³ nitrogen) with simultaneous cleavage of the bond between the carbon and the amino nitrogen. The unimolecular decomposition reaction of formamidine was first studied by Andrés et al. [133, 134] and other researchers [126,127]. Their activation energy for this reaction was 333 kJ mol⁻¹ at the HF/4-31G level of theory. In their study [133, 134], the formamidine structure, reactant minima, had C_s symmetry. However, our vibrational analysis, as reported in Chapter

3, shows that this structure is not the minimum and it has one imaginary frequency, and the nearly planar C_1 structure has the minimum energy. Kaushik et al. [126,127] studied the unimolecular decomposition of the *Z* isomer of formamidine. They found the energy barrier for this reaction is $328.1 \text{ kJ mol}^{-1}$ at AM1 [126] and $279.6 \text{ kJ mol}^{-1}$ at B3LYP/6-31G(d) [127] levels of theory.

To my knowledge, no computational studies of the water-mediated or solution-phase decomposition of formamidine that have been reported. Proton transfer in an aqueous medium is important in all biological systems. Therefore, one of the goals of this work was to study this reaction in both the gas phase and in solution. Solvent molecules can act not only as a solvent but also as a catalyst to stabilize the transition state of the reacting system, thus, lowering the activation energy. For these reasons, we have performed a comprehensive study of the unimolecular and the water-mono- and di-hydrated decomposition of formamidine in the gas phase. We have also paid attention to studying this reaction in solution using the PCM solvent model. The results of this work have provided a model to study the intramolecular and intermolecular hydrogen rearrangement of cytosine and to choose the computationally less expensive level of theory which will give a reliable energetic result. These investigations also will aid us to understand the hydrogen transfer mechanism and the interactions in solution.

1. 8 References

- (1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 171, 964–967.
- (2) Calladine, C. R.; Drew, H. R.; Luisi, B. F.; Travers, A. A. *Understanding DNA*, Elsevier Academic Press, London, **2004**.
- (3) Blackburn, G. M.; Gait, M. J. “*Nucleic Acids in Chemistry and Biology*”, Oxford University Press, NewYork, **1996**.
- (4) Ulbricht, T. L. V. “*Purines Pyrimidines and Nucleotides and the Chemistry of Nucleic Acids*”, Pergamon Press, New York, **1964**.
- (5) Neidle, S. “*Oxford Hand Book of Nucleic Acid Structure*”; Oxford University Press Inc.: New York, **1999**.
- (6) Townsend, L. B. “*Chemistry of Nucleosides and Nucleotides*”, Volume 1, Plenum Press, NewYork, **1988**.
- (7) Löwdin, P. O. *Rev. Mod. Phys.* **1963**, 35, 724.
- (8) Löwdin, P. O. *Adv. Quantum Chem.* **1965**, 2, 213.
- (9) Bertran, J.; Oliva, A.; Rodríguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, 120, 8159–8167.
- (10) Florián, J.; Leszczyński, J. *J. Am. Chem. Soc.* **1996**, 118, 3010–3017.
- (11) Florián, J.; Hroudá, V.; Hobza, P. *J. Am. Chem. Soc.* **1994**, 116, 1457–1460.
- (12) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem.* **1996**, 100, 1965–1974.

- (13) Scheiner, S.; Kern, C. W. *J. Am. Chem. Soc.* **1979**, 101, 15, 4081.
- (14) Hroudá, V.; Florián, J.; Polášek, M.; Hobza, P. *J. Phys. Chem.* **1994**, 98, 4742–4747.
- (15) Hroudá, V.; Florián, J.; Hobza, P. *J. Phys. Chem.* **1993**, 97, 1542.
- (16) Hobza, P.; Šponer, J.; Polášek, M. *J. Am. Chem. Soc.* **1995**, 117, 792–798.
- (17) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Comp. Chem.* **1996**, 17, 841–850.
- (18) Smets, J.; Jalbout, A. F.; Adamowicz, L. *Chem. Phys. Lett.* **2001**, 342, 342–346.
- (19) Danilov, V. I.; Leś, A.; Alderfer, J. L. *Polish J. Chem.* **2001**, 75, 1039–1049.
- (20) Harrison, C. B.; O’Neil, L. L.; Wiest, O. *J. Phys. Chem. A* **2005**, 109, 7001–7012.
- (21) Tureček, F.; Yao, C. *J. Phys. Chem. A* **2003**, 107, 9221–9231.
- (22) Zhang, H.; Liang, Q.; Xia, Y.; Zhao, M.; Ji, Y.; Song, C.; Liu, X.; Zhang, B. *Int. J. Quantum Chem.* **2007**, 107, 240–246.
- (23) Szczesniak, M.; Leszczyński, J.; Person, W. B. *J. Am. Chem. Soc.* **1992**, 114, 2731–2733.
- (24) Doerfler, W.; Böhm, P. *DNA Methylation: Basic Mechanism*. Springer, New York, **2006**.
- (25) Fülcher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, 117, 2089–2095.
- (26) Florián, J.; Baumruk, V.; Leszczyński, J. *J. Phys. Chem.* **1996**, 100, 5578–5589.
- (27) Ji, Y. J.; Xia, Y. Y.; Zhao, M. W.; Huang, B. D.; Li, F. *J. Mol. Struct: THEOCHEM* **2005**, 723, 123–129.
- (28) Hou, X.-J.; Nguyen, M. T. *Chem. Phys.* **2005**, 310, 1–9.
- (29) Wysokiński, R.; Bienko, D. C.; Michalska, D.; Zeegers-Huyskens, T. *Chem. Phys.* **2005**, 315, 17–26.

- (30) Dong, W.; Wang, H.; Ren, X.; Shan, Y.; Ge, Q. *J. Solution Chem.* **2007**, 36, 549–561.
- (31) Chandra, A. K.; Nguyen, M. T.; Zeegers-Huyskens, T. *J. Mol. Struct.* **2000**, 519, 1–11.
- (32) Chandra, A. K.; Michalska, D.; Wysokinsky, R.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **2004**, 108, 9593–9600.
- (33) Morpurgo, S.; Bossa, M.; Morpurgo, G. O. *Advances In Quantum Chemistry* **2000**, 36, 169.
- (34) Broo, A.; Holmén, A. *J. Phys. Chem. A* **1997**, 101, 3589–3600.
- (35) Hunter, K. C.; Rutledge, L. R.; Wetmore, S. D. *J. Phys. Chem. A* **2005**, 109, 9554–9562.
- (36) Hunter, K. C.; Wetmore, S. D. *Chem. Phys. Lett.* **2006**, 422, 500–506.
- (37) Monajjemi, M.; Ghiasi, R.; Ketabi, S.; Passdar, H.; Mollaamin, F. *J. Chemical Research* **2004**, 11–18.
- (38) Burda, J.; Šponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem. B* **1997**, 101, 9670–9677.
- (39) Monajjemi, M.; Ghiasi, R.; Abedi, A. *Theoretical Inorganic Chemistry* **2005**, 50, 435–441.
- (40) Šponer, J.; Burda, J. V.; Sabat, M.; Leszczyński, J.; Hobza, P. *J. Phys. Chem. A* **1998**, 102, 5951–5957.
- (41) Burda, J.; Šponer, J.; Hobza, P. *J. Phys. Chem.* **1996**, 100, 7250–7255.
- (42) Prado, M. A. S.; Garcia, E.; Martins, J. B. L. *Chem. Phys. Lett.* **2006**, 418, 264–267.

- (43) Fuentes-Cabrera, M.; Sumpter, B. G.; Šponer, J. E.; Šponer, J. *J. Phys. Chem. B* **2007**, 111, 870–879.
- (44) Kistler, K. A.; Matsika, S. *J. Phys. Chem. A* **2007**, 111, 2650–2661.
- (45) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, 110, 8319–8330.
- (46) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1989**, 111, 2308–2310.
- (47) Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, 98, 6338.
- (48) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Comput. Chem.* **1996**, 17, 841–850.
- (49) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, 98, 3161–3164.
- (50) Estrin, D. A.; Paglieri, L.; Corongiu, G. *J. Phys. Chem.* **1994**, 98, 5653–5660.
- (51) Scanlan, M. J.; Hillier, I. H. *J. Am. Chem. Soc.* **1984**, 106, 3737–3745.
- (52) Civcir, P. Ü. *J. Mol. Struct. (THEOCHEM)* **2000**, 532, 157–169.
- (53) Gould, I. R.; Green, D. V. S.; Young, P.; Hillier, I. H. *J. Org. Chem.* **1992**, 57, 4434–4437.
- (54) Alemà, C. *Chem. Phys.* **2000**, 253, 13–19.
- (55) Sambrano, J. R.; Souza, A. R.; Queralt, J. J.; Andrés, J. *Chem. Phys. Lett.* **2000**, 317, 437–443.
- (56) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, 118, 6811–6821.
- (57) Leś, A.; Adamowicz, L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, 93, 4001–4005.

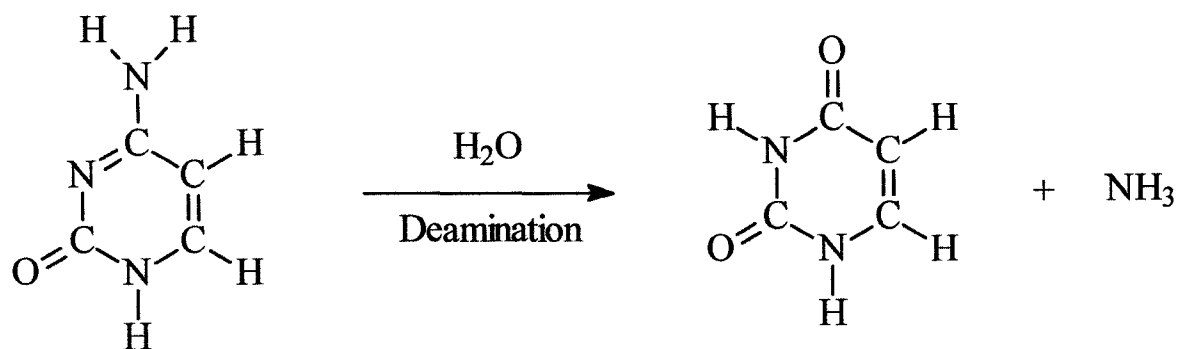
- (58) Person, W.B.; Szczepaniak, K.; Szczesniak, M.; Kwiatkowski, J. S.; Hernandez, L.; Czerminski, R. *J. Mol. Struc: THEOCHEM* **1989**, 194, 239.
- (59) Goddard, J. D.; Mezey, P. G; Csizmadia, I. J. *Theor. Chim. Acta.* **1975**, 39, 1.
- (60) Leszczyński, J. *J. Phys. Chem.* **1992**, 96, 1649–1653.
- (61) Laudo, M. D.; Whittleton, S. R.; Wetmore, S. D. *J. Phys. Chem. A* **2003**, 107, 10406–10413.
- (62) Hunter, K. C.; Millen, A. L.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, 111, 1858–1871.
- (63) Duchesne, J. “*Physico-Chemical Properties of Nucleic Acids*”, Academic Press: Volume 3, New York, **1973**.
- (64) Devlin, T. M. “*Textbook of Biochemistry with Clinical Correlations*”, Wiley-Liss, New York, **1997**.
- (65) Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, 106, 506–555.
- (66) Notari, R. E.; Chin, M. L.; Cardoni, A. *J. Pharmaceutical Science* **1970**, 59, 28.
- (67) Shapiro, R.; Klein, R. *Biochemistry* **1966**, 5, 2358.
- (68) Shapiro, R.; Servis, R. E.; Wecher, M. *J. Am. Chem. Soc.* **1970**, 92, 422.
- (69) Shapiro, R.; Klein, R. *Biochemistry* **1967**, 6, 3576.
- (70) Shapiro, R.; DiFate, V.; Welcher, M. *J. Am. Chem. Soc.* **1974**, 906–912.
- (71) Lindahl, T.; Nyberg, B. *Biochemistry* **1974**, 13, 3405–3410.
- (72) Chen, H.; Shaw, B. R. *Biochemistry* **1994**, 33, 4121–4129.
- (73) Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. *Biochemistry* **1990**, 29, 2532–2537.

- (74) Voss, K. O.; Roos, K. P.; Nonay, R. L.; Dovichi, N. J. *Anal. Chem.* **1998**, 70, 3818–3823.
- (75) Yao, L.; Li, Y.; Wu, Y.; Liu, A.; Yan, H. *Biochemistry* **2005**, 44, 5940–5947.
- (76) Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, Jr. C. W. *Biochemistry* **1997**, 36, 4768–4774.
- (77) Carlow, D. C.; Smith, A. A.; Yang, C. C.; Short, S. A.; Wolfenden, R. *Biochemistry* **1995**, 34, 4220–4224.
- (78) Snider, M. J.; Reinhardt, L.; Wolfenden, R.; Cleland, W. W. *Biochemistry* **2002**, 41, 415–421.
- (79) Matsubara, T.; Ishikura, M.; Aida, M. *J. Chem. Inf. Model.* **2006**, 46, 1276–1285.
- (80) Glaser, R.; Man-Shick, S. *J. Am. Chem. Soc.* **1996**, 118, 10942–10943.
- (81) Glaser, R.; Rayat, S.; Lewis, M.; Man-Shick, S.; Meyer, S. *J. Am. Chem. Soc.* **1999**, 121, 6108–6119.
- (82) Rayat, S.; Qian, M.; Glaser, R. *Chem. Res. Toxicol.* **2005**, 18, 1211–1218.
- (83) Glaser, R.; Wu, H.; Lewis, M. *J. Am. Chem. Soc.* **2005**, 127, 7346–7358.
- (84) Samaranayake, M.; Bujnicki, J. M.; Carpenter, M.; Bhagwat, A. S. *Chem. Rev.* **2006**, 106, 700–719.
- (85) Merchant, K.; Chen, H.; Gonzalez, T. C.; Keefer, L. K.; Shaw, B. R. *Chem. Res. Toxicol.* **1996**, 9, 891–896.
- (86) Malia, S. A.; Basu, A. K. *Chem. Res. Toxicol.* **1994**, 7, 823–828.

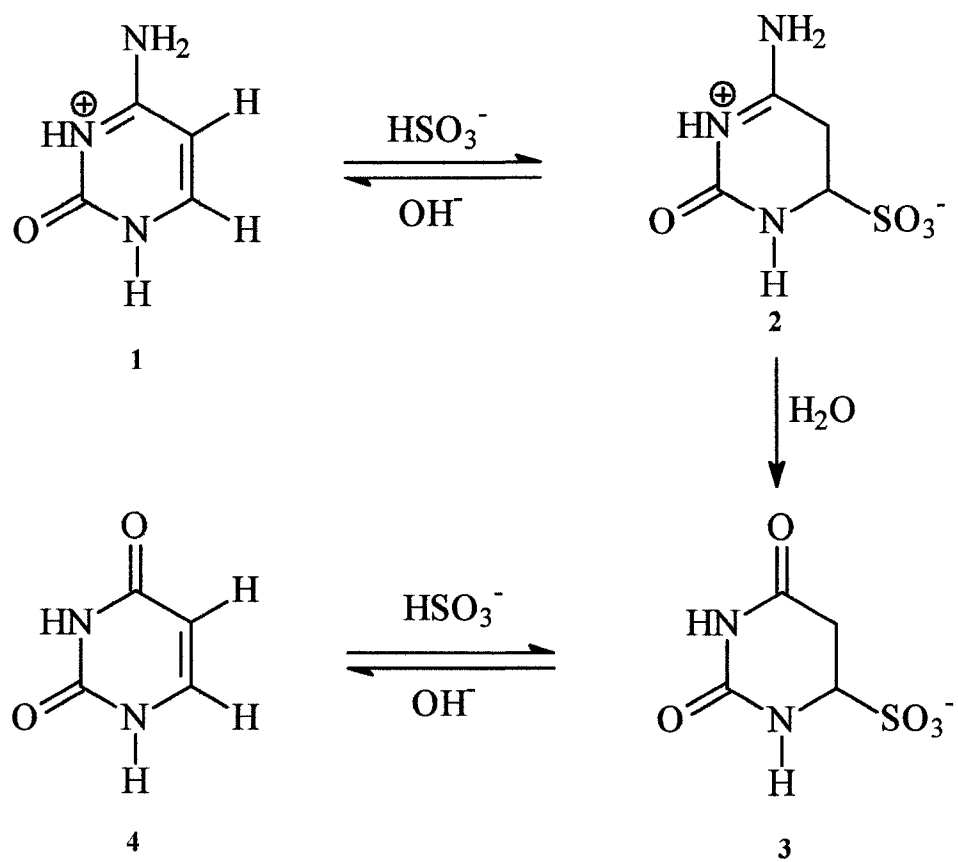
- (87) Sohail, A.; Hayes, C. S.; Divvela, P.; Setlow, P.; Bhajwat, A. S. *Biochemistry* **2002**, *41*, 11325–11330.
- (88) Sponer, J. E.; Miguel, P. J.; Rodriguez-Santiago, L.; Erxleben, A.; Krumm, M.; Sodupe, M.; Sponer, J.; Lippert, B. *Angew. Chem. Int. Ed.* **2004**, *43*, 5396–5399.
- (89) Sherer, G.; Limbach, H-H. *J. Am. Chem. Soc.* **1989**, *111*, 5946–5947.
- (90) Sherer, G.; Limbach, H-H. *J. Am. Chem. Soc.* **1994**, *116*, 1230.
- (91) Rumpel, H.; Limbach, H-H. *J. Am. Chem. Soc.* **1989**, *111*, 5429.
- (92) Belasco, J. G.; Albery, W. J.; Knowles, J. R. *Biochemistry* **1986**, *25*, 2552.
- (93) Poirier, R. A.; Wang, Y.; Westaway, K. C. *J. Am. Chem. Soc.* **1994**, *116*, 2526.
- (94) Glad, S. S.; Jenser, F. *J. Am. Chem. Soc.* **1994**, *116*, 9302.
- (95) Bendict, H.; Shenderovich, I. G.; Malkina, O. L.; Malkin, V. G.; Denisov, G. S.; Goluber, N. S.; Limbech, H-H. *J. Am. Chem. Soc.* **2000**, *122*, 1979–1988.
- (96) Liu, Q.; Hoffman, R. *J. Am. Chem. Soc.* **1995**, *117*, 10108–10112.
- (97) Lim, J-H.; Lee, E. K.; Kim, Y. *J. Phys. Chem. A* **1997**, *101*, 2233–2239.
- (98) Zhanpeisov, N.; Cox, Jr, W. W.; Leszczynski, J. *J. Phys. Chem. A* **1999**, *103*, 4564–4571.
- (99) Clementt, E.; Mehl, J.; Von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508.
- (100) Grout, R. J. J. *The chemistry of Amidines and Imidates*. Edited by Patai, S. John Wiley & Son, NewYork, **1975**.
- (101) Shriner, R. L.; Neamann, F. W. *Chem. Rev.* **1944**, *35*, 351.
- (102) Flinn, C.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2003**, *107*, 11171.

- (103) Almatarnah, M. H.; Flinn, C. G.; Poirier, R. A. *Can. J. Chem.* **2005**, *83*, 2082–2090.
- (104) Yamabe, T.; Yamashita, K.; Kaminoyama, M.; Koizumi, M.; Tachibana, A. *J. Phys. Chem.* **1984**, *88*, 1459.
- (105) Nguyen, K. A.; Gordon, M. S.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 1596.
- (106) Zielinski, T. J.; Poirier, R. A. *J. Comp. Chem.* **1984**, *5*, 5, 466.
- (107) Poirier, R. A.; Majlessi, D.; Zielinski, T. J. *J. Comp. Chem.* **1986**, *7*, 4, 464.
- (108) Poirier, R. A.; Yu, D.; P. Surjan, R. *Can. J. Chem.* **1991**, *69*, 1589.
- (109) Kim, Y.; Lim, S.; Kim, H-H.; Kim, Y. *J. Phys. Chem. A* **1999**, *103*, 617.
- (110) Pecul, M.; Leszczynski, J.; Sadlej, J. *J. Phys. Chem. A* **2000**, *104*, 8105.
- (111) Wiberg, K. B.; Rablen, P. R. *J. Am. Chem. Soc.* **1995**, *117*, 2201.
- (112) Czernek, J. *J. Phys. Chem. A* **2003**, *107*, 3952.
- (113) Wang, X-C.; Nichols, J.; Feyereisen, M.; Gutowski, M.; Boatz, J.; Haymet, A. D. J.; Simons, J. *J. Phys. Chem.* **1991**, *95*, 10419.
- (114) Dziekonski, P.; Sokalski, W. A.; Podolyan, Y.; Leszczynski, J. *Chem. Phys. Lett.* **2003**, *367*, 367.
- (115) Podolyan, Y.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 12103.
- (116) Zhanpeisov, N. U.; Leszczynski, J. *J. Phys. Chem. A* **1999**, *103*, 8317–8327.
- (117) Shimon, L.; Glusker, J. P.; Bock, C. W. *J. Phys. Chem.* **1996**, *100*, 2957.
- (118) Kim, Y.; Lim, S.; Kim, Y. *J. Phys. Chem. A* **1999**, *103*, 6632.
- (119) Zhang, Q.; Bell, R.; Truong, T. N. *J. Phys. Chem.* **1995**, *99*, 592.
- (120) Nagaoka, M.; Okuno, Y.; Yamabe, T. *Can. J. Chem.* **1992**, *70*, 377.

- (121) Nagaoka, M.; Okuno, Y.; Yamabe, T. *J. Am. Chem. Soc.* **1991**, 113, 769.
- (122) Tortajada, J.; Leon, E.; Luna, A.; Mo, O.; Yanez, M. *J. Phys. Chem.* **1994**, 98, 12919.
- (123) Kim, Y.; *J. Phys. Chem. A* **1998**, 102, 3025.
- (124) Bell, R. L.; Truong, T. N. *J. Phys. Chem. A* **1997**, 101, 7802.
- (125) Yamashita, K.; Kaminoyama, M.; Yamabe, T.; Fukui, K. *Theo. Chim. Acta.* **1981**, 60, 303.
- (126) Kaushik, R.; Rastogi, R. C.; Ray, N. K. *J. Indian Chem. Soc.* **1993**, 70, 289–293.
- (127) Kaushik, R.; Rastogi, R. C.; Ray, N. K. *Indian J. Chem. Sec. A* **1996**, 35, 629–632.
- (128) Nagaoka, M.; Okuno, Y.; Yamabe, T.; Fukui, K. *Can. J. Chem.* **1992**, 70, 377–387.
- (129) Zielinski, T. J.; Peterson, M. R.; Csizmadia, I. G.; Rein, R. *J. Comput. Chem.* **1982**, 3, 62.
- (130) Radom, L.; Hehre, W. J.; Pople, J. *J. Am. Chem. Soc.* **1971**, 93, 289.
- (131) Filleux, M. L.; Naulet, N.; Doric, J. P.; Martin, G. J.; Pornet, J.; Miginiac, L. *Tetrahedron Lett.* **1974**, 1435.
- (132) Yamabe, T.; Yamashita, K.; Kaminoyama, M.; Koizumi, M.; Tachibana, A.; Fukui, K. *J. Phys. Chem.* **1984**, 88, 1459.
- (133) Andres, J.; Krechl, J.; Carda, M.; Silla, E. *In. J. Quantum. Chem.* **1991**, XL, 127.
- (134) Andres, J.; Beltran, A.; Carda, M.; Krechl, J.; Monterde, J.; Silla, E. *J. Mol. Struct.: THEOCHEM* **1992**, 254, 465.

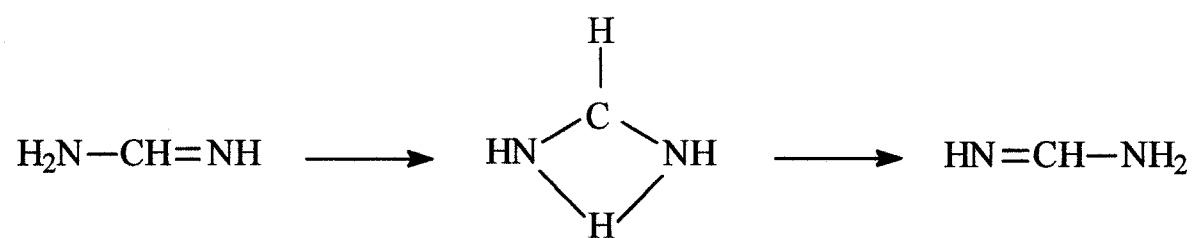


Scheme 1.1: Deamination reaction of cytosine.

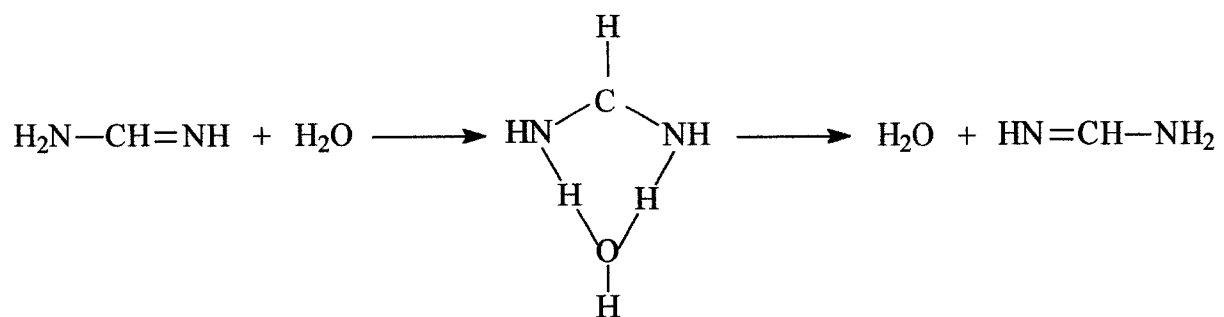


Scheme 1.2:

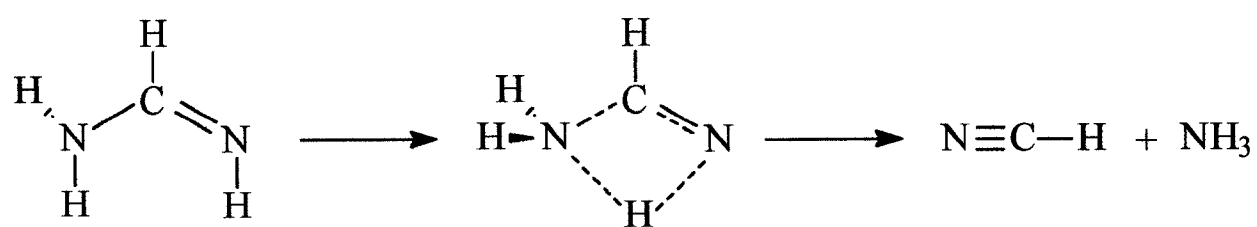
Deamination reaction of cytosine catalysed by sulfite.



Scheme 1.3: Intramolecular hydrogen-transfer rearrangement of formamidine.



Scheme 1.4: Water-mediated intramolecular hydrogen-transfer rearrangement of formamidine.



Scheme 1.5: The unimolecular decomposition reaction of formamidine.

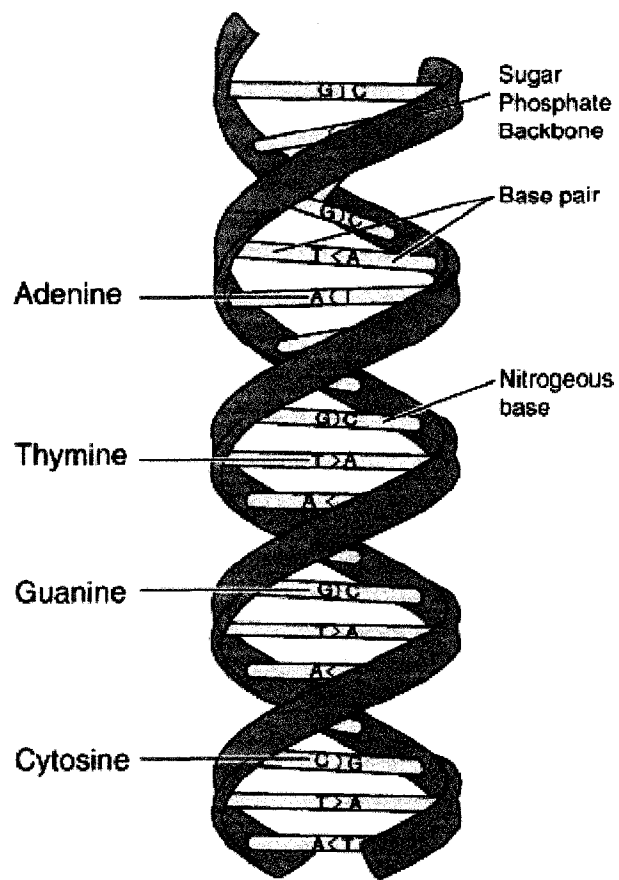
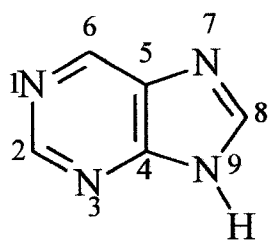
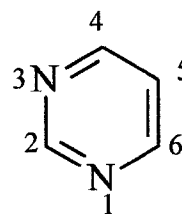


Figure 1.1: The Watson-Crick double helix of DNA where the base pairs are hydrogen bonded³.

³<http://www.askfactmaster.com/Image/DNA-structure-and-bases.png>



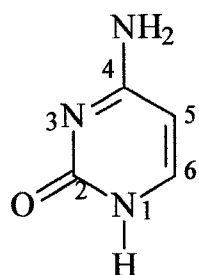
Purine



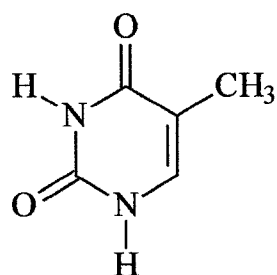
Pyrimidine

Figure 1.2:

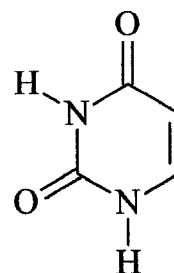
General structures of pyrimidine and purine.



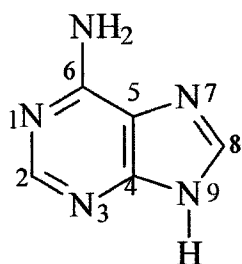
Cytosine



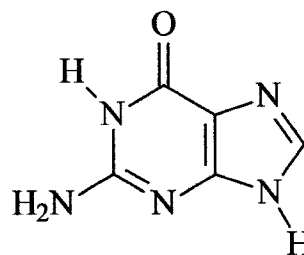
Thymine



Uracil

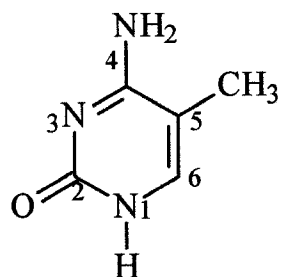


Adenine

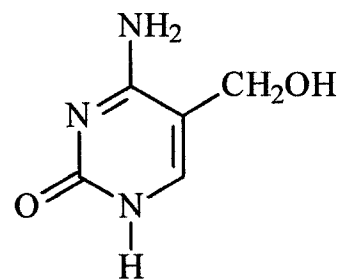


Guanine

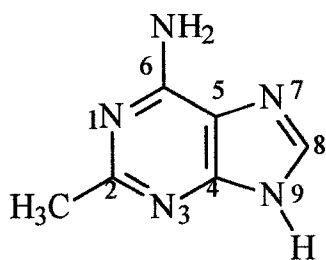
Figure 1.3: Structures of the major pyrimidine and purine bases of nucleic acids.



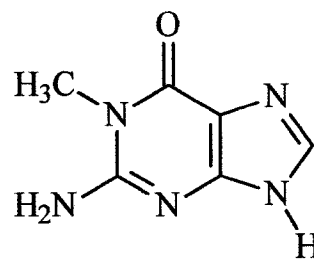
5-Methylcytosine



5-Hydroxymethylcytosine



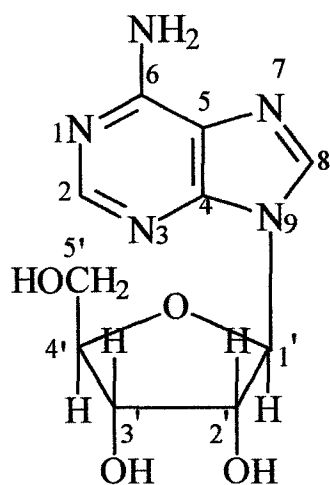
2-Methyladenine



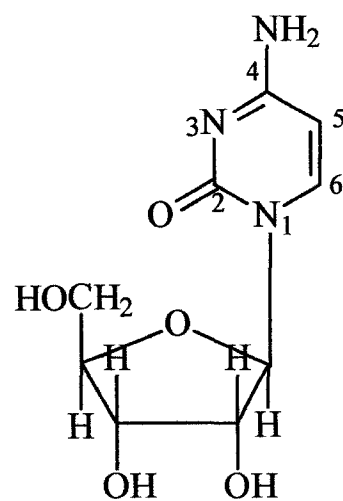
1-Methylguanine

Figure 1.4:

Structures of the minor pyrimidine and purine bases of nucleic acids.

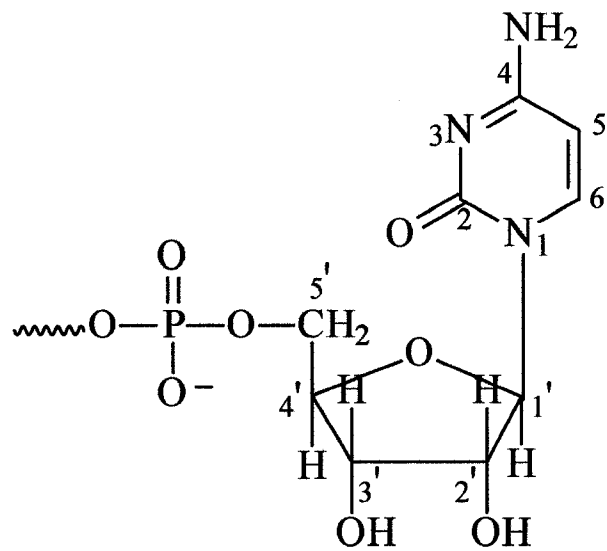


Purine nucleoside (Adenosine)



Pyrimidine nucleoside (Cytidine)

Figure 1.5: Structures of the ribonucleosides.



Cytidine 5'-monophosphate

Figure 1.6:

Structure of cytidine 5'-monophosphate.

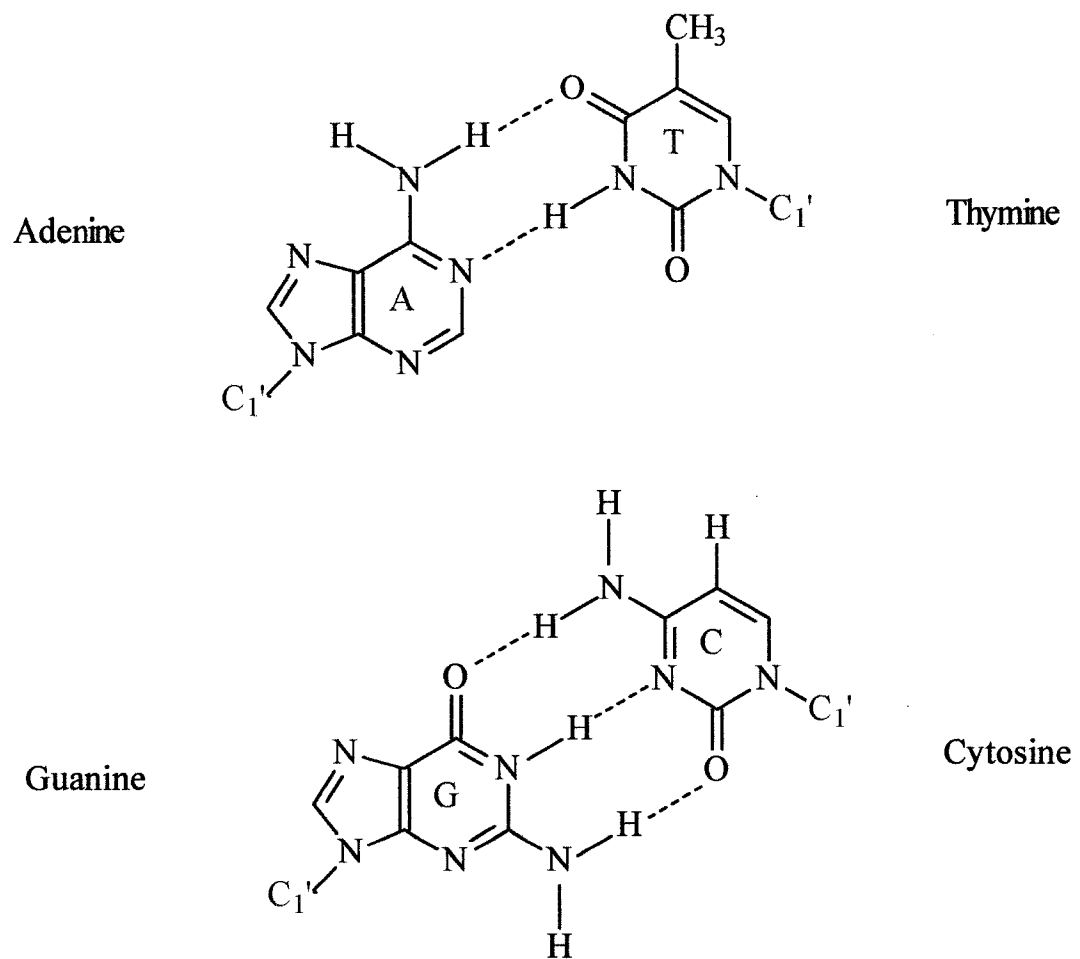
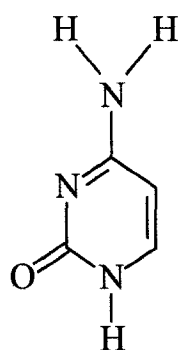
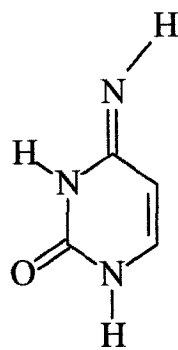


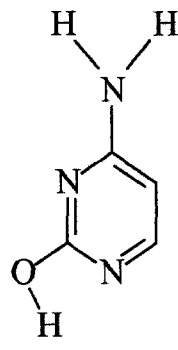
Figure 1.7: Watson-Crick base pairing for A•T and G•C.



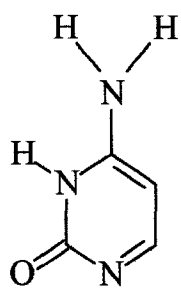
C1



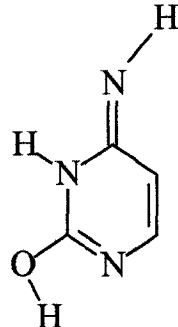
C2



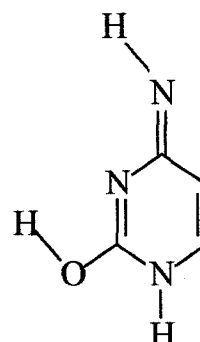
C3



C4

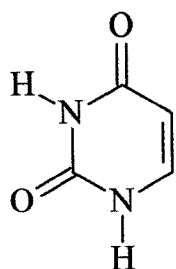


C5

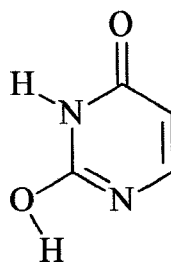


C6

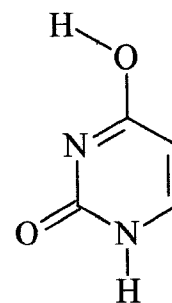
Figure 1.8: Six tautomers of cytosine.



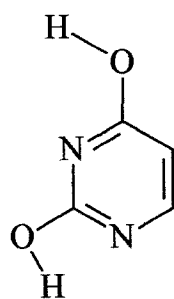
U1



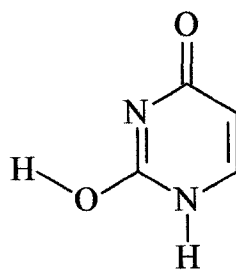
U2



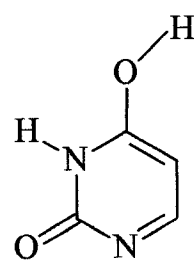
U3



U4



U5



U6

Figure 1.9: Six tautomers of uracil .

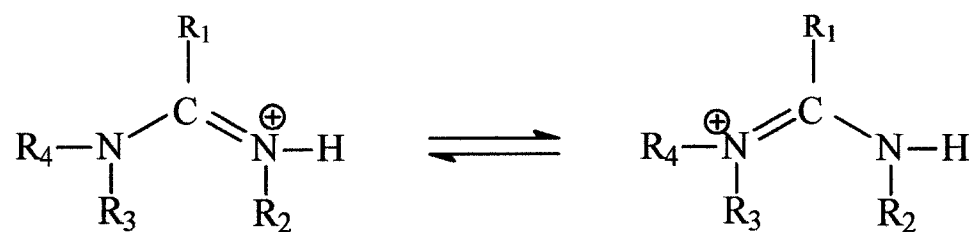


Figure 1.10: Resonance structures of amidinium ions.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 The Schrödinger Equation

The main goal of computational chemists is focusing on finding approximate solutions of the time-independent, non-relativistic Schrödinger equation [1,2].

$$\hat{H}\Psi_i(\vec{r},\vec{R}) = E_i\Psi_i(\vec{r},\vec{R}) \quad (1)$$

where, \vec{r} is the electronic coordinate and \vec{R} is the nuclear coordinate. \hat{H} is the Hamiltonian operator (or simply, the Hamiltonian) for a molecular system consisting of M nuclei and N electrons. Ψ_i is the wavefunction of the i 'th state of the system. It is a function of the positions of all fundamental particles (electrons and nuclei) in the system. Ψ depends on $4N$ variables, three spatial variables (coordinates) and one spin variable for each of the N electrons along with $3M$ spatial coordinates of the nuclei. E_i represents the total energy of the system. The

Hamiltonian, \hat{H} , contains all the terms that contribute to the total energy of the system which can be written in the atomic units for N electrons and M nuclei as:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (2)$$

Here, A and B run over the M nuclei while i and j run over the N electrons in the system. The first two terms describe the kinetic energy of the electrons and nuclei, respectively. The Laplacian operators ∇_q^2 ($q = i$ and A) have the form

$$\nabla_q^2 = \frac{\partial^2}{\partial X_q^2} + \frac{\partial^2}{\partial Y_q^2} + \frac{\partial^2}{\partial Z_q^2} \quad (3)$$

The third term in Eq. (2) represents the attractive electrostatic interaction between the nuclei and the electrons (Coulomb attraction). The last two terms represent the repulsive potential due to the electron-electron and nucleus-nucleus interaction, respectively. R_{pq} (and similarly R_{qp}) is the distance between the particles p and q , i. e., $r_{ij} = |r_i - r_j|$, $r_{iA} = |r_i - R_A|$ and $R_{AB} = |R_A - R_B|$.

2.2 Born-Oppenheimer Approximation

In 1927, Born and Oppenheimer [4] showed that, to a very good approximation, the motions of the nuclei and the electrons could be considered separately. This is the famous Born-Oppenheimer approximation, in which the electrons are considered to be moving in the field of fixed nuclei. Thus, the complete Hamiltonian given in Eq. (2) reduces to the electronic Hamiltonian,

$$H_{elec} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} \quad (4)$$

The Schrödinger equation using the electronic Hamiltonian operator is

$$\hat{H}_{elec} \psi_{elec} = E_{elec} \psi_{elec} \quad (5)$$

where, ψ_{elec} depends on the electron coordinates and parametrically on the nuclear coordinates. The total energy E_{tot} is then the sum of E_{elec} and the constant nuclear repulsion term (V_{nn}), i. e.,

$$E_{tot} = E_{elec} + V_{nn} \quad (6)$$

The eigenvalue of the electronic Schrödinger equation is called the ‘electronic energy’. From now on, we will only consider the electronic problem in Eq. (5) and Eq. (6) and the subscript ‘elec’ will be dropped. The total energy $E_{tot}(R)$ is a function of the nuclear coordinates, which

provides a potential energy surface (PES). Thus, the concept of PES is a result of the Born-Oppenheimer approximation [1, 5].

In bra-ket notation, we let $|\Psi\rangle$ represents the ket vector for wavefunction Ψ , and $\langle\Psi| = \Psi^*$ represents the bra vector for wavefunction Ψ . By using this notation, the expectation value of the Hamiltonian operator is given by:

$$\int \Psi^* \hat{H} \Psi d\tau = \langle\Psi|\hat{H}|\Psi\rangle = E \quad (7)$$

where Ψ is normalized. The star in Ψ^* indicates the complex-conjugate of Ψ .

2.3 The Variational Principle

The variation theorem states that the energy determined from any approximate (trial) wavefunction, will be always greater than or equal to the energy for the exact wavefunction (E_{exact}) [1,2,5], i. e.,

$$E[\Psi] = \langle\Psi_{trial}|\hat{H}|\Psi_{trial}\rangle \geq E_{exact} \quad (8)$$

where $E[\Psi]$ is the energy functional.

2.4 Slater Determinants

In the simplest approximation for the many-electron systems, the wavefunction should satisfy the following conditions [1,3,5]:

1. The wavefunction must be antisymmetric with respect to the interchanges of the spatial and spin coordinates of any two electrons.

$$\Psi(r_1, r_2, \dots, r_i, r_j, \dots, r_N) = -\Psi(r_1, r_2, \dots, r_j, r_i, \dots, r_N) \quad (9)$$

2. Satisfy Pauli exclusion principle (no two electrons can occupy the same state).
3. Electrons must be indistinguishable, i. e., if the coordinates of any two electrons are switched, the probability must stay the same.

The simplest wavefunction that satisfies these conditions is the wavefunction which is written as a Slater determinant. The Slater determinant in terms of spin orbitals, χ_i , can be written as:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_N(2) \\ \vdots & \vdots & \dots & \vdots \\ \chi_1(N) & \chi_2(N) & \dots & \chi_N(N) \end{vmatrix} \quad (10)$$

where $(N!)^{-1/2}$ is a normalization factor. This Slater determinant has N electrons and the spin orbitals are composed of a product of spatial (molecular) orbitals, and one of the two spin functions. The spin orbitals can be written as:

$$\chi(x) = \psi(r) \sigma(\omega) \quad , \quad \sigma = \alpha, \beta \quad (11)$$

where $\psi(r)$ is the molecular orbital (MO), $\alpha(\omega)$ is a spin up function and $\beta(\omega)$ is a spin down function. Slater determinants are often written in short hand notation, where only the diagonal elements are given:

$$\Psi = \left| \chi_1(x_1) \chi_2(x_2) \dots \chi_N(x_N) \right\rangle \quad (12)$$

which can be further shortened to

$$\Psi = \left| \psi_1 \bar{\psi}_1 \dots \psi_N \bar{\psi}_N \right\rangle \quad (13)$$

where ψ_1 and $\bar{\psi}_1$ stand for the MO's with spin up and spin down electrons, respectively.

2.5 The Hartree-Fock Approximation

This section will show the final results without any derivations, an excellent source for an in-depth explanation of many aspects of the Hartree-Fock (HF) approximation is the book by Szabo and Ostlund, 1982 [1]. HF theory is the simplest approximation for solving the time-independent Schrödinger equation [1-3,5].

The HF energy is given by

$$E_{HF} = \langle \Psi | \hat{H} | \Psi \rangle = \sum_a^N (a | \hat{h} | a) + \frac{1}{2} \sum_a^N \sum_b^N [aa|bb] - [ab|ba] \quad (14)$$

in chemist's notation, where \hat{H} is the full electronic Hamiltonian. (a and b) are spin orbitals. Eq.

(14) can be written in physicist's notation as:

$$E_{HF} = \sum_a^N \langle a | \hat{h} | a \rangle + \frac{1}{2} \sum_a^N \sum_b^n \langle ab | ab \rangle \quad (15)$$

where the first term in Eq. (14) is the one-electron integral shown below

$$[a | \hat{h} | a] = \langle a | \hat{h} | a \rangle = \int \chi_a^*(1) \left\{ -\frac{1}{2} \nabla_1^2 - \sum_A^M \frac{Z_A}{r_{1A}} \right\} \chi_a(1) dx_1 = h_{aa} \quad (16)$$

which defines the contribution from the kinetic energy and the electron-nucleus attraction. The second term in Eq (14) is defined as

$$[aa | bb] = \iint |\chi_a(1)|^2 \frac{1}{r_{12}} |\chi_b(2)|^2 dx_1 dx_2 = \langle ab | ab \rangle = J_{ab} \quad (17)$$

$$[ab | ba] = \iint \chi_a^*(1) \chi_b(1) \frac{1}{r_{12}} \chi_b^*(2) \chi_a(2) dx_1 dx_2 = \langle ab | ba \rangle = K_{ab} \quad (18)$$

Eq. (17) and Eq. (18) represent the Coulomb (J_{ab}) and exchange (K_{ab}) integrals, respectively, which indicate the interaction between two electrons. Eq. (14) can then be written as:

$$E_{HF} = 2 \sum_a^N h_{aa} + \sum_a^N \sum_b^N 2J_{ab} - K_{ab} \quad (19)$$

According to the variation theorem, the best wavefunction is the one which gives the lowest possible energy. For N-electron systems, minimizing the energy with respect to the spin orbitals, expanding and integrating over spin orbitals, this leads to Hatree-Fock equations,

$$\hat{f}_i \psi_i = \varepsilon_i \psi_i, i = 1, 2, \dots, N \quad (20)$$

where, ε_i are the eigenvalues of the Fock operator \hat{f}_i . Roothaan introduced a set of K known spatial basis functions $\{\phi_\mu, \mu = 1, 2, \dots, K\}$ to obtain a numerical solution for molecules by expanding the unknown MO's as a linear expansion. Molecular orbitals are expressed as a linear combination of a set of basis functions as:

$$\psi_i(r) = \sum_{\mu=1}^K C_{\mu i} \phi_\mu(r) \quad (21)$$

where $C_{\mu i}$ are the molecular orbital expansion coefficients. The set of K functions $\phi_\mu(r)$ is called the 'basis set'. Basis functions will be discussed in the next section. The Roothaan equation can be written as a single matrix equation

$$F C = \varepsilon S C \quad (22)$$

where C is a $K \times K$ square matrix which represents the molecular orbital coefficients. F is the Fock matrix which can be written as:

$$F_{\mu\nu} = \int \phi_{\mu}^*(1) f(1) \phi_{\nu}(1) dr_1 \quad (23)$$

S is the overlap matrix which can be written as:

$$S_{\mu\nu} = \int \phi_{\mu}^*(1) \phi_{\nu}(1) dr_1 \quad (24)$$

The Fock operator is an effective one-electron operator defined as:

$$\hat{f}(i) = -\frac{1}{2} \nabla_i^2 - \sum_A^M \frac{Z_A}{r_{iA}} + V^{HF}(i) = h(i) + V^{HF}(i) \quad (25)$$

Here, $V^{HF}(i)$ is the Hartree-Fock potential. It is the average repulsive potential experienced by the i 'th electron due to all the other remaining electrons. The $V^{HF}(i)$ can be written as:

$$V^{HF}(i) = \sum_b^N (\hat{J}_b(1) - \hat{K}_b(1)) \quad (26)$$

\hat{J}_b is the Coulomb operator which is defined as

$$\hat{J}_b(1) \chi_a(1) = \left[\int \chi_b^*(2) r_{12}^{-1} \chi_b(2) \right] \chi_a(1) = \langle \chi_b(2) | r_{12}^{-1} | \chi_b(2) \rangle \chi_a(1) \quad (27)$$

The exchange operator \hat{K}_b is defined as

$$\hat{K}_b(1) \chi_a(1) = \left[\int \chi_b^*(2) r_{12}^{-1} \chi_a(2) dx_2 \right] \chi_b(1) = \langle \chi_b(2) | r_{12}^{-1} | \chi_a(2) \rangle \chi_b(1) \quad (28)$$

The Coulomb operator has a classical interpretation which corresponds to the classical Coulombic interaction. The exchange operator defines the electron exchange energy and has no classical interpretation. The procedure for solving the Roothaan equation iteratively is called the Self-Consistent-Field (SCF).

2. 6 Basis set Expansions

The construction in Eq. (21) is known as the linear combination of atomic orbitals (LCAO) approach, where the molecular orbital (MO) coefficients $C_{\mu i}$ are calculated variationally through the SCF procedure.

Gaussian type functions (or orbitals) (GTF or GTO) have simpler integral calculations than Slater type orbitals (STO). This is due to the fact that the product of any two Gaussian is also a Gaussian. The Gaussian orbital has the general form

$$g_{l,m,n}(\alpha_i) = N_i (X - A_X)^l (Y - A_Y)^m (Z - A_Z)^n \exp(\alpha_i r_A^2) \quad (29)$$

where, (X, Y, Z) are electron coordinates; (A_X, A_Y, A_Z) are nuclear coordinates (origin of Gaussian); N_i is a normalization factor; and α_i is the Gaussian orbital exponent. r_A^2 is given by:

$$r_A^2 = (X - A_X)^2 + (Y - A_Y)^2 + (Z - A_Z)^2 \quad (30)$$

In general, basis functions, ϕ^i s, are a linear combination of GTF,

$$\phi_{\mu}^t = \sum_{r=1}^{N_g} d_r g_r^t(\alpha_r) \quad (31)$$

where N_g is typically, 1 to 6; d_r is a contraction coefficient; and $g_r^t(\alpha_r)$ is the Gaussian primitive function with exponent α_r of a given type $t(s, p, d, \dots)$. Examples of Gaussian basis sets are: 3-21G, 6-31G(d), 6-31+G(d,p),...etc. The smallest basis sets are called minimal basis sets, which are composed of one basis function for each occupied AO. Appendix (A) shows an example of the number of basis functions for cytosine and cytosine anion.

2. 7 The Post-Hartree-Fock Methods

The HF method is a single determinant wavefunction. It has one main drawback which is neglecting the correlation energy. The exchange occurs between electrons with the same spin only, which is correctly represented by HF, while electrons with opposite spins are not correlated. The difference between the experimental ‘exact’ and the Hartree-Fock limit energies is called the correlation energy (E_{corr}).

$$E_{corr} = E_0 - E_{HFL} \quad (32)$$

The Hartree-Fock limit energy, E_{HFL} , is obtained in the limit that the basis set approaches completeness. The E_{HFL} would still be higher than the exact non-relativistic energy, thus the correlation energy is always negative. Computing the correlation energy is a very important

problem in quantum chemistry due to the fact that neglecting E_{corr} can lead to large deviations from experimental results. To include the electron correlation, we can write the wavefunction as a multideterminant wavefunction. There are many methods that exist to compute E_{corr} , each with its own strengths and weaknesses, collectively called Post-Hartree-Fock methods. Some of these methods are:

1. Configuration Interaction (CI), Multi-Configuration SCF (MCSCF), and Quadratic Configuration Interaction (QCI).
2. Møller-Plesset perturbation theory of order n (MP n).
3. Coupled Cluster (CC).

These methods provide more accurate results than HF for treating electron correlation, however, they are very costly in computer time. Appendix (B) shows a comparison between some of these methods, particularly, the cost (the computer resources consumed) of an SCF calculation. There is an extensive literature on the post-Hartree-Fock methods, for a detailed overview, see Ref. [7]. I should point out that none of these methods are exact. All these methods are *ab initio* quantum methods, and do not rely on empirical parameters.

2. 7. 1 Size-Extensivity and Size-Consistency

Size-extensivity and size-consistency are important concepts in electronic structure theory. They are used somewhat interchangeably in the literature. To define these terms, let us consider two non-interacting H_2 molecules which dissociate into two H_2 molecules [1,5,6,8]. A method is size-extensive, if the energy of the two non-interacting H_2 molecules are calculated

simultaneously, then the energy should be exactly twice the one calculated for the individual molecule, i. e.,

$$E(\text{H}_2\text{--H}_2) = 2E(\text{H}_2)$$

We will find that as distance ($\text{H}_2\text{--H}_2$) $\rightarrow \infty$, the energy difference will equal zero. Here, we introduced the size-extensivity concept, where the potential energy curve should be calculated by scaling linearly with the number of particles. In the case when the non-interacting H_2 molecules are getting close together, they must be correctly described. Size-consistency describes the proper behaviour of the wavefunction in dissociation processes. Thus, a method is not size-consistent when it yields poor dissociation energies. HF, MPn, and full CI are size extensive.

2. 7. 2 Configuration Interaction

The HF wavefunction is a single determinant referred to as a single configuration treatment. Different configurations can be created by ‘exciting’ one or more electrons from occupied to virtual orbitals [1]. These configurations can be mixed together to obtain a better approximation to the wavefunction:

$$\Psi_{CI} = C_0 \Psi_0 + \sum_{ar} C_a^r |\Psi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} C_{ab}^{rs} |\Psi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} C_{abc}^{rst} |\Psi_{abc}^{rst}\rangle + \dots \quad (33)$$

By including all the excitations, Eq. (33) is then the full CI wavefunction. The CI coefficients

$(C_0, C_a^r, C_{ab}^{rs}, \dots)$ can be found via the variation theorem. Full CI is only feasible for small molecules and modest basis sets. Truncated CI methods limit the types of excitations that can occur. The most common CI methods are CISD, QCISD and QCISD(T), where ‘SD(T)’ stands for single, double, and triple excitations. The last method is among the most accurate methods, however, it is also the most expensive computational wavefunction based method, it scales as $O(N^7)$ (N being the number of basis functions). CI is a variational method. Full CI is size-extensive, but truncated CI is not, see Appendix (B).

2. 7. 3 Møller-Plesset Perturbation Theory

In perturbation theory (PT) , the Hamiltonian is divided into two parts:

$$\hat{H} = \hat{H}_0 + \lambda \hat{V} \quad (34)$$

where \hat{H}_0 is the unperturbed Hamiltonian operator and \hat{V} is the perturbation operator with λ being an arbitrary parameter. The energy are expanded as a power series which can be written as:

$$E = E^{(0)} + E^{(1)} + E^{(2)} + \dots \quad (35)$$

where $E^{(n)}$ is the n th-order energy correction. Expansion to first order gives the Hartree-Fock energy,

$$\begin{aligned}
E^{(0)} + E^{(1)} &= \langle \Psi^{(0)} | \hat{H}^{(0)} | \Psi^{(0)} \rangle + \langle \Psi^{(0)} | \hat{V}^{(1)} | \Psi^{(0)} \rangle \\
&= \langle \Psi^{(0)} | \hat{H} | \Psi^{(0)} \rangle = E_{HF}
\end{aligned} \tag{36}$$

where $\Psi^{(0)}$ and $E^{(0)}$ are the HF wavefunction and energy, respectively. Møller and Plesset have applied perturbation theory by defining \hat{V} which treats the electron correlation effects, usually to second (MP2), third (MP3), or fourth (MP4) order. The second-order Møller-Plesset perturbation theory (MP2) is given by:

$$E_{MP2} = \sum_{a < b}^{occ} \sum_{r < s}^{vir} \frac{\langle \Psi_0 | \hat{V} | \Psi_{ab}^{rs} \rangle \langle \Psi_{ab}^{rs} | \hat{V} | \Psi_0 \rangle}{(\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s)} = \sum_{\substack{a < b \\ r < s}} \frac{|\langle ab || rs \rangle|^2}{(\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s)} \tag{37}$$

while MP2 scales with fifth power of the system size, MP4 scales as $O(N^7)$. MPn energy is size-extensive but not variational.

2.8 Density Functional Theory (DFT)

In DFT, the ground-state wavefunction, molecular energy, and all other molecular electronic properties are calculated by the ground-state electron density $\rho(r)$ [9–15],

$$E = E_0[\rho] \quad (38)$$

where the energy depends on the function $\rho(r)$, where the square brackets denote a functional relation. $\rho(r)$ is a function of three variables (x, y, z). Unlike the wavefunction, the electron density is an observable and can be measured experimentally by X-ray diffraction. In 1964, Hohenberg and Kohn (HK) [16], proved two theorems which were critical to establishing DFT as a quantum chemistry methodology. Kohn and Sham (KS) in 1965 [17], devised a practical method for finding ground-state densities ρ_0 and ground-state energies E_0 . The energy density functional has the form

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + V_{ee}[\rho] + \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (39)$$

where,

$$T_s = \langle \Psi | \hat{T} | \Psi \rangle = -\frac{1}{2} \sum_{i \in occ} \langle \Psi_i^{KS} | \nabla^2 | \Psi_i^{KS} \rangle \quad (40)$$

Ψ is the Slater determinant. T_s is the kinetic energy for the reference system of non-interacting electrons (uniform electron gas or Jellium), where the Kohn-Sham orbitals, Ψ^{KS} , are included in the non-interacting electron density. The density for the reference systems is written as:

$$\rho_S = \sum_{i=1}^N \left| \Psi_i^{KS} \right|^2 \quad (41)$$

$V_{ne} [\rho]$ represents the classical potential energy in terms of the density, which is given by

$$V_{ne} [\rho] = \sum_A^M \int \frac{Z_A}{r - r_A} \rho(r) dr \quad (42)$$

$V_{ee} [\rho]$ represents the classical expression for the electrostatic electron-electron repulsion functional given by:

$$V_{ee} [\rho] = \frac{1}{2} \iint \frac{\rho(r_1) \rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 \quad (43)$$

The last two terms in Eq. (39) are corrections to the kinetic energy of the fictitious reference system and all the non-classical corrections to the electron-electron repulsion energy, respectively. These two terms are combined together in one term as:

$$E_{xc} [\rho] = \Delta T [\rho] + \Delta V_{ee} [\rho] \quad (44)$$

where $E_{xc} [\rho]$ is the exchange-correlation energy functional. Eq. (39) can then be written as:

$$E [\rho] = T_s [\rho] + V_{ne} [\rho] + V_{ee} [\rho] + E_{xc} [\rho] \quad (45)$$

The key to accurate KS DFT calculation of molecular properties is to get a good approximation

to E_{XC} . Based on the fictitious system of non-interacting electrons and after minimizing $E[\rho]$, the Kohn-Sham equation is then written as:

$$\hat{h}_i^{KS}(1) \Psi_i^{KS}(1) = \varepsilon_i^{KS} \Psi_i^{KS}(1) \quad (46)$$

where,

$$\hat{h}^{KS}(1) = -\frac{1}{2} \nabla^2 + \sum_A^M \frac{Z_A}{r_{1A}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + \hat{V}_{XC}(1) \quad (47)$$

where the exchange-correlation potential, \hat{V}_{XC} , is found as the functional derivative of the E_{XC} energy. \hat{h}^{KS} is the Kohn-Sham Hamiltonian. HK showed that if the density varies slowly with position, then the E_{XC} can be written as:

$$E_{XC}[\rho] = \int \rho(r) \varepsilon_{XC}[\rho] dr \quad (48)$$

where the integral is over all space, and dr stands for dx , dy , and dz . The functional E_{XC} is written as the sum of an exchange-energy functional and a correlation-energy functional as:

$$\varepsilon_{XC}(\rho) = \varepsilon_X(\rho) + \varepsilon_C(\rho) \quad (49)$$

DFT has a number of functionals (X : exchange functional and C : correlation functional), such as:

1. Local density approximation (LDA), e.g., Vosko-Wilk-Nusair (VWN) functional [18].

2. Generalized gradient approximation (GGA), e. g., Becke88 (B88) [19], Perdew86 [20], PW91 [21], and Lee-Yang-Parr (LYP) functionals [22].
3. Hybrid functionals, e. g., B3LYP, B3PW91, and B1B96.

Becke3-Lee-Yang-Parr (B3LYP) [22–24] is the most popular hybrid density functional that is used and yields very good structural and thermochemical properties.

$$E_{XC}^{B3LYP}[\rho] = (1 - C_1)E_X^{LSDA}[\rho] + C_1E_K^{HF}[\rho] + C_2E_X^{B88}[\rho] + C_3E_C^{LYP}[\rho] + (1 - C_3)E_C^{VWN}[\rho] \quad (50)$$

where ($C_1 = 0.20$, $C_2 = 0.72$, $C_3 = 0.83$) are parameters which were chosen to give a good fit to experimental molecular atomization energies. LSDA stands for local spin density approximation (exchange functional). Any exchange functional can be combined with any correlation functional. For example, the notation BLYP/6-31G(d) denotes: A density functional done with the Becke1988 exchange functional in conjunction with the Lee-Yang-Parr correlation functional with KS orbitals expanded in a 6-31G(d) basis set.

2.9 Computational Methodology

Computational chemistry uses calculations to study systems of interest to a wide variety of scientists, i.e., chemists, biologists, biochemists, etc. Over the last two decades, powerful molecular modelling tools have been developed, which are capable of accurately predicting structures, energetics, reactivities and other properties of molecules. These improvements have come about largely due to the increase of computer speed and the design of efficient quantum chemical algorithms. Several works have reported that *ab initio* structures are generally in good agreement with experimental data, which makes *ab initio* methods a useful tool for investigating molecular structures which are not known experimentally.

Most of the electronic structure calculations in this thesis were performed with either Gaussian98 (G98) [25] or Gaussian03 (G03)[26] quantum chemical packages. The MUNgauss [27] computational program was used for most of the geometry optimization at HF/6-31G(d) for the decomposition of formamidine and the deamination of cytosine with H₂O and OH⁻. The geometries of all reactants, transition states, intermediates, and products were fully optimized at *ab initio* methods using the restricted Hartree-Fock RHF level of theory, and MP2 [28] using 6-31G(d) [29] and 6-31+G(d) [30] basis sets. We have also used density functional theory by performing B3LYP calculations [22-24] to investigate the change in the potential energy surface (PES) compared with the *ab initio* levels of theory.

Single point energies were also calculated using Gaussian-n theories, B3LYP/6-31+G(d)

(for the deamination reaction of cytosine with H_2O and OH^-), and MP2/GTMP2Large levels of theory. The Gaussian-n theories, which have been used to study the decomposition reaction of formamidine, are G1 [31, 32], G2 [33], G2MP2 [34], G3 [35], G3MP2 [36], G3B3 [37], and G3MP2B3 [37] levels of theory. Single-point energies required for G3MP2 level of theory was obtained using MP2(FULL)/6-31G(d) optimized geometry. It is known that Gaussian-n theories produce very accurate energies, which in most cases are close to chemical accuracy. However, these levels of theory are too computationally expensive to be used for large systems. These calculations are practical only for very small systems. Thus, we were looking for a more economical level of theory to give reliable energetic results. For this reason, we have chosen to study the decomposition of formamidine as a simple model, because of its small size.

We first focussed on a simple amidine compound, formamidine, which will be considered as a simple model for the study of cytosine. Through comparison of Gaussian-n theories, the most efficient level of theory for this system was determined. From our results on the decomposition of formamidine, we found that the activation energies and thermodynamic properties calculated using Gaussian-n theories all agreed to within 10 kJ mol^{-1} , which is within the reported error of these levels in the literature. On the basis of these results, we chose the G3MP2 level of theory [36] which is the cheaper variant of the G3 level of theory [35]. The G3MP2 calculations will be discussed in the next section. Appendix (C) shows the CPU time for performing the geometry optimization of structures involved in the decomposition of formamidine with one and two water molecules. These results show that the G3MP2 level of

theory is the least computationally expensive method for providing reliable energetics.

To get the whole picture for all the mechanisms proposed in this thesis, we have carried out intrinsic reaction coordinates (IRC) analysis [38–40] to characterize all the optimized transition states. In order to identify the true minimum, we optimized the last IRC points which are closest to the next local minimum in order to positively identify the reactant, intermediate and product to which each transition state is connected. A frequency analysis was performed for each stationary point in order to confirm whether this point represented a minimum on the PES (no imaginary frequencies) or a transition state (one imaginary frequency).

While gas-phase studies are of fundamental interest, the effect of the environment such as solvation, must also be considered. We have investigated the unimolecular decomposition reaction of formamidine with the polarizable continuum model (PCM) with different choices of cavity, for example, KLAMT, PAULING, and UFF. The solvent models will be discussed in section (2.12), which will give a brief introduction about the solvation energies as well as an example. For more detail on using and execution of jobs using G03, see the Gaussian03 keywords at http://www.gaussian.com/g_ur/keywords.htm.

2. 10 G3MP2 Theory

The Gaussian-*n* theories are computationally expensive for calculating molecular energy to high accuracy. The major goal of these methods is to accurately calculate, $\pm 2 \text{ kcal mol}^{-1}$ ($\pm 8.4 \text{ kJ mol}^{-1}$), properties such as: atomization energies; electron affinities, ionization potentials; and proton affinities. Calculations using Gaussian-*n* theories are practical only for very small systems. The G3MP2 level of theory consists of several calculations which are combined together to give reliable energetic results. This theory is based on optimized geometries at MP2(FULL)/6-31G(d). The other energy calculations are done with a frozen core (fc) approximation in the treatment of electron correlation.

The average mean absolute deviation (MAD) from experiment of G3 theory are 1.01 and $0.94 \text{ kcal mol}^{-1}$ (4.2 and 3.9 kJ mol^{-1}) for the 299 energies and the subset of 148 neutral enthalpies, respectively [35]. The mean absolute deviation for G3MP2 level of theory are 1.30 and $1.18 \text{ kcal mol}^{-1}$ (5.4 kJ mol^{-1} and 4.9 kJ mol^{-1}) for the 299 energies and the 148 neutral enthalpies, respectively [36]. The G3MP2 level of theory is analogous to the G2MP2 theory [34], which is computationally less expensive than G2 theory [33].

The G3MP2 total energy at 0 K is defined in the following equation:

$$\begin{aligned} E_0(\text{G3MP2}) = & E[\text{QCISD}(T, \text{FC}) / 6-31\text{G}(d) // \text{MP2}(\text{FULL})6-31\text{G}(d)] \\ & + \Delta E(\text{MP2}) \\ & + \Delta E(\text{HLC}) \\ & + E(\text{ZPE}) \\ & + \Delta E(\text{SO}) \end{aligned} \quad (51)$$

where,

ΔE (MP2) is given by:

$$\Delta E(MP2) = E[MP2(FC) / G3MP2Large // MP2(FULL) / 6-31G(d)] - E[MP2(FC) / 6-31G(d) // MP2(FULL) / 6-31G(d)] \quad (52)$$

ΔE (HLC) is the higher level correction which is added to compensate for the remaining deficiencies in the energy calculations. ΔE (HLC) is given by:

$$\Delta E(HLC) = -An_{\beta} - B(n_{\alpha} - n_{\beta}) \quad (53)$$

The n_{α} and n_{β} are the number of α and β valence electrons, respectively, with $n_{\alpha} \geq n_{\beta}$. The A and B values are chosen to give the smallest average absolute deviation for the test set of experimental energies. For G3MP2 theory, the A and B values are:

- (1) $A = 9.279$ mhartrees, and $B = 4.471$ mhartrees (for molecules).
- (2) $A = 9.345$ mhartrees, and $B = 2.021$ mhartrees (for atoms, including atomic ions).

$E(ZPE)$, the zero-point energy correction, is obtained from HF/6-31G(d) frequencies and then scaled by a factor of 0.8929.

$$E(ZPE) = 0.8929 \times [ZPE(HF / 6-31G(d))] \quad (54)$$

ΔE (SO), a spin-orbit correction, is mainly taken from experimental results, which is included only for atomic species. Appendix (D) shows an example of a G1, G2, and G3MP2 calculation, in which the calculations are done both ‘manually’ (by doing the calculations at each level of theory separately) and using a G3MP2 command directly as implemented in Gaussian03.

2. 11 Solvation Models

2. 11. 1 Polarizable Continuum Model (PCM)

Many chemical and biological reactions occur in water, where polar and ionic processes are much more favourable than in the gas phase. The free energy of solvation, a fundamental quantity, is the free energy difference associated with the transfer of a solute X from the gas phase to a given solvent Y [41]. It describes the interaction of the solute with the solvent in which it is dissolved. The use of continuum models can lead to very interesting results since they allow a detailed description of the changes caused by the solute molecule due to its interaction with the solvent. Generally, the interaction between solute and the solvent depends on the geometry of the solute. Thus, the computation of the free energy of solvation should involve an average over solute geometries. It is a reasonable assumption to perform a single point energy calculation for most solutes using either the optimized gas phase geometry or the optimized liquid phase geometry, and this is the procedure employed in most cases. However, it is dangerous in the case of mechanistic studies to do single-point energy calculations using gas phase optimized structures, in which you assume that the mechanism in the gas phase and in the solution are the same, which is not always true, as shown from my calculations in Chapter 3.

The polarizable continuum model (PCM), initially devised by Tomasi and co-workers [42–47], and COSMO (conductor-like-screening model) were applied to calculate the free

energies of solvation of the unimolecular decomposition reaction of formamidine. The cavity in the PCM model is created via a series of overlapping spheres, and COSMO-RS (conductor-like-screening model for real solvents) [48-51] request a conductor PCM calculation (CPCM) using atomic radii and other parameters as suggested by Klamt [52, 53] for his models, which are provided as a module in Gaussian03. The PCM is one of the many successful solvation models, in which the solute interacts with the solvent represented by a dielectric continuum model.

In the Gaussian03 quantum package, the cavity can be selected in a number of different ways. The PCM defines the cavities as envelopes of spheres centered on atoms or atomic groups. The solvent for the PCM calculation is water, which is a solvent default in Gaussian03. The choice of cavities in PCM is important because the computed energies and properties depend on the cavity size. By default, the program builds up the cavity using the United Atom (UA0) model, i.e., by putting a sphere around each solute heavy atom; hydrogen atoms are enclosed in the sphere of the atom to which they are bonded. There are three UA models available which were not suitable for our reaction due to the fact that our systems involve proton transfer reactions. The cavity can be extensively modified in the PCM input section³ by putting spheres around specified hydrogens, changing sphere parameters and topology of the general cavity, adding extra spheres to the default cavity, and so on. A number of cavity models have been suggested. In this thesis, the UFF, PAULING, BONDI, and KLAMT cavities were used to evaluate the aqueous solvation effect on the decomposition of formamidine using PCM.

³ http://www.gaussian.com/g_ur/k_scrf.htm

The UFF use radii from the UFF force field in which hydrogens have individual spheres. For the PAULING and BONDI cavities, each solute atom and group is assigned van der Waals values obtained from Pauling [54] or Bondi [55] atomic radii.

2. 11. 2 PCM Calculations

This section shows how we calculate the free energies of reaction and the free energies of activation using the HF *ab initio* method.

(I) Using Hartree-Fock theory

1- For the gas phase

The free energy of reaction is given by:

$$\Delta G_{\text{gas}} = G_{\text{gas}}^{\text{total}}(\text{P}) - G_{\text{gas}}^{\text{total}}(\text{R}) \quad (55)$$

and the free energy of activation (act) is given by:

$$\Delta G_{\text{gas}}^{\text{act}} = G_{\text{gas}}^{\text{total}}(\text{TS}) - G_{\text{gas}}^{\text{total}}(\text{R}) \quad (56)$$

The R, TS, and P stand for reactant, transition state, and product, respectively. The free energy for any structure can be expressed by:

$$G_{\text{gas}}^{\text{total}}(\text{X}) = E_{\text{HF}}(\text{X}) + \Delta\Delta G_{\text{gas}}^{\text{thermal}} \quad (57)$$

while the thermal correction is given in the output of Gaussian and can be calculated by:

$$\Delta\Delta G_{\text{gas}}^{\text{thermal}} = G_{\text{gas}} - E_{\text{gas}} \quad (58)$$

using the free energies given in the Gaussian output, where,

$\Delta\Delta G_{\text{gas}}^{\text{thermal}}$: Thermal correction to Gibbs free energy (from Gaussian output).

$G_{\text{gas}}^{\text{total}}(X)$: Sum of electronic energy and thermal free energies

X: R, TS, or P.

Appendix (E) shows an example of a PCM calculation as well as the input keywords for optimization and single point calculations, along with the output summary for an example from the decomposition reaction of formamidine.

2. For the solution phase

(A) Using single point calculations:

The free energy of reaction is given by:

$$\Delta G_{\text{soln}} = G_{\text{soln}}^{\text{total}}(\text{P}) - G_{\text{soln}}^{\text{total}}(\text{R}) \quad (59)$$

and the free energy of activation is given by:

$$\Delta G_{\text{soln}}^{\text{act}} = G_{\text{soln}}^{\text{total}}(\text{TS}) - G_{\text{soln}}^{\text{total}}(\text{R}) \quad (60)$$

where,

$$G_{\text{soln}}^{\text{total}}(\text{X}) = G_{\text{soln}}(\text{X}) + \Delta\Delta G_{\text{gas}}^{\text{thermal}} \quad (61)$$

(B) Using optimized structures:

The free energy of reaction and activation are given by Eq. (59) and Eq. (60), respectively.

where,

$$G_{\text{soln}}^{\text{total}}(\text{X}) = G_{\text{soln}}(\text{X}) + \Delta\Delta G_{\text{soln}}^{\text{thermal}} \quad (62)$$

Calculations with MP2 level of theory will follow the same equations as in HF calculations.

Note that the free energy of solvation at MP2 is the value saved in the Gaussian archive.

2. 12 **References**

- (1) Szabo, A.; Ostlund, N. "*Modern Quantum Chemistry*", Dover Publication, INC., New York, **1996**.
- (2) Levine, I. N. "*Quantum Chemistry*", Prentice-Hall, Inc., New Jersey, **2000**.
- (3) Pilar, F. L. "*Elementary Quantum Chemistry*", McGraw-Hill Publishing Company, New York, **1990**.
- (4) Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *84*, 457.
- (5) Cramer, C. "*Computational Chemistry*", John Wiley and Sons, Ltd., Chichester, England, **2004**.
- (6) Hinchliffe, A. "*Computational Quantum Chemistry*", John Wiley and Sons, Chichester, England, **1988**.
- (7) Bartlett, R. J.; Stanton, J. F. "*Applications of Post-Hartree-Fock Methods: A Tutorial*" *Rev. Comput. Chem.* **1994**, *5*, 65–169.
- (8) Surján, P. R. "*Second Quantized Approach to Quantum Chemistry*", Springer-Verlag, Berlin, Germany, **1989**.
- (9) Fiolhais, C.; Nogueira, F.; Marques, M. "*A Primer in Density Functional Theory*", Springer, Berlin, Germany, **2003**.
- (10) Gross, E. K. U.; Dreizler, R. M. "*Density Functional Theory*", Plenum Press, New York, **1995**.

- (11) Labanowski, J. K.; Andzelm, J. W. "*Density Functional Methods in Chemistry*", Springer-Verlag, New York, **1991**.
- (12) Holthausen, M. C.; Koch, W. "*A Chemist's Guide to Density Functional Theory*", Wiley-VCH, Weinheim, Germany, **2001**.
- (13) Sahni, V. "*Quantal Density Functional Theory*", Springer, Berlin, Germany, **2004**.
- (14) Chong, D. P. "*Recent Advances in Density Functional Methods*", Vol. 1, World Scientific, Ltd., Singapore, **1997**.
- (15) Schleyer, M. F. "*Encyclopedia of Computational Chemistry*", John Wiley and Sons, New York, **1998**.
- (16) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, 136, 864.
- (17) Kohn, W.; Sham, L. *J. Phys. Rev.* **1965**, 140, A1133.
- (18) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58, 1200.
- (19) Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098.
- (20) Perdew, J. P. *Phys. Rev. B* **1986**, 33, 8822.
- (21) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, 48, 4978.
- (22) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785.
- (23) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648.
- (24) Becke, A. D. *J. Chem. Phys.* **1996**, 104, 1040.
- (25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman J. R.; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich,

S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; *Gaussian98*, Revision A.11.3, Gaussian, Inc., Pittsburgh PA, **2002**.

(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr. T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.;

Gonzalez, C.; and Pople, J. A. *Gaussian03*, Revision B.05, Gaussian, Inc.: Pittsburgh PA, **2003**.

(27) Poirier, R. A.; *MUNgauss* (Fortran 90 version) [computer program]. Chemistry Department, Memorial University of Newfoundland, St. John's, Newfoundland, A1B 3X7 Canada. With contributions from Brooker, M. ; Bungay, S. D.; El-Sherbiny, A.; Gosse, T.; Keefe, D.; Pye, C. C.; Reid, D.; Shaw, M.; Wang, Y.; Xidos, J.; (b) Colonna, F.; Jolly, L-H.; Poirier, R. A.; Angyan, J.; Jansen, G.; *Comput. Phys. Commun.* **1994**, 81, 293-317.

(28) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, 46, 618–622.

(29) (a) Petersson, G. A.; Al-Laham, M. A. *J. Chem. Phys.* **1991**, 94, 6081.

(b) Petersson, G. A.; Bennett, A.; Gensfeldt, T. G.; Al-Laham, M. A. *J. Chem. Phys.* **1988**, 89, 2193.

(30) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta.* **1973**, 28, 213–222

(31) Pople, J. A.; Head-Gordon, M. ; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, 90, 5622.

(32) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1990**, 93, 2537.

(33) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, 94, 7221.

(34) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, 98, 1293.

(35) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem Phys.* **1998**, 109, 7764–7776.

(36) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V., Pople, J. A. *J. Chem. Phys.*

1999, 110, 4703–4709.

(37) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, 110, 7650.

(38) Fukui, K. *Acc. Chem. Res.* **1981**, 14, 363.

(39) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1989**, 90, 2154.

(40) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, 94, 5523.

(41) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, 108, 6532.

(42) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, 55, 117.

(43) Miertus, S.; Tomasi, J. *J. Chem. Phys.* **1982**, 65, 239.

(44) Cossi, M.; Barone, R.; Cammi, R.; Tomasi, J. *J. Chem. Phys. Lett.* **1996**, 255, 327.

(45) Cossi, M.; Barone, V.; Robb, M. A. *J. Chem. Phys.* **1999**, 111, 5295.

(46) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, 106, 5151.

(47) (a) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, 117, 43–54. (b)

Barone, V.; Impropa, R.; Rega, N. *Theor. Chem. Acc.* **2004**, 111, 237–245.

(48) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(49) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, 103, 9312–9320.

(50) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, 102, 1995–2001.

(51) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, 24, 669–681.

(52) Klamt, A. *J. Phys. Chem.* **1995**, 99, 2224.

(53) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, 102, 5074.

(54) Weast, R. C., Ed.; *Handbook of Chemistry and Physics*; Chemical Rubber: Cleveland, OH, **1981**.

(55) Bondi, A. J. *Chem. Phys.* **1964**, 68, 441–451.

CHAPTER 3

Ab initio Study of the Decomposition Reaction of Formamidine ⁴

3.1 Introduction

Amidine compounds are of interest because of their medical and biochemical importance, having many biological and pharmaceutical uses [1]. Formamidine is the simplest of the amidine class and has been found to exhibit antibiotic, antifungal, and anaesthetic properties. Amidines play a significant role in the biosynthesis of other biologically important compounds, such as purines, imidazoles, and the catabolism of histidine [1]. Due to its small size, formamidine has been a prime target of both experimental and theoretical investigations.

⁴Almatarneh, M. H.; Flinn, C.; Poirier, R. *Can. J. Chem.* **2005**, *83*, 2082-2090.

In addition to serving as a model for hydrogen transfer reactions in bases of nucleic acids, formamidine has been extensively studied theoretically since it forms hydrogen bonded complexes with itself and other molecules such as water.

The intramolecular 1,3-sigmatropic hydrogen rearrangement and intermolecular hydrogen transfers have been studied theoretically for various formamidine systems [2–9] including formamide-formamidine [10–17] and formamidine-formic acid [18–19]. Since most hydrogen transfers occur in aqueous solution, one must consider the role of water molecules in hydrogen transfer. Water can act not only as a solvent but also as a catalyst by both donating and accepting a hydrogen to promote long range hydrogen transfer (water-mediated pathway).

Tautomerisation by intramolecular hydrogen transfer involving formamidine and its complexes with one, two and three water molecules was studied by Zhang et al. [20] using *ab initio* and density functional methods. They reported the importance of hydrogen-bonded water molecules which lowered the barrier to tautomerisation. Nguyen et al. [6] investigated double hydrogen transfer in formamidine dimers using *ab initio* methods. The barrier for the prototropic tautomerization in these systems is reduced by about 84 kJ mol⁻¹ at Hartree-Fock level of theory when the hydrogen transfer is mediated by a water molecule. The solvent effect on the potential energy surface for double hydrogen transfer in formamidine dimers has also been studied [21]. Solvent effects on the potential energy surface have also been investigated for formamidine hydrogen bonded to a water molecule using the chemical molecular dynamic

simulation method [2].

Since biological activity depends greatly on the molecular conformation, Tortajada et al. [23] have investigated the relative stabilities of the *E(trans)* and *Z(cis)* conformers of formamidine and the relative stabilities of the complexes of both conformers with different monocations. They found that the (*E*) conformer of formamidine is 7.5 kJ mol^{-1} more stable than the *Z* conformer at both G2MP2 and G2 levels of theory. Dynamics and kinetic isotope effects for tautomerization of the formamidine monohydrated complex using direct semiempirical dynamic calculation and G2 level of theory was studied by Kim [24]. Primary and solvent kinetic isotope effects in the water-mediated tautomerisation of formamidine using an *ab initio* direct dynamic study was investigated by Bel et al. [25].

A number of reactions of formamidine have been considered to be model reactions for more complex molecules with structural similarities to formamidine, such as, single hydrogen transfer in bases of nucleic acids [6, 8, 26] as well as double hydrogen transfer reactions. Recently, the deamination of the more stable *E*-isomer of formamidine with OH^- , H_2O , and H_3O^+ to yield formamide has been studied by Flinn et al. [26]. This investigation was undertaken as a precursor to a study of the deamination of the DNA base cytosine. The *E (trans)* isomer of formamidine forms part of the cytosine structure, notably in the region of cytosine where the deamination of cytosine to uracil can take place. Cytosine is the most unstable of the DNA bases, deaminating spontaneously to uracil with an activation energy of $117 \pm 4 \text{ kJ mol}^{-1}$ [27].

The mechanism for the deamination of formamidine was investigated using high level *ab initio* methods. Flinn et al. found that deamination with OH^- yields a tetrahedral intermediate with a much lower activation energy barrier than for the reaction with H_2O and H_3O^+ using G2 theory.

Amidine is also known to undergo decomposition at elevated temperatures, yielding hydrogen cyanide and ammonia. Some experimental results suggest that the decomposition of amidine is an unimolecular process [28]. The unimolecular decomposition mechanism of formamidine to give ammonia and hydrogen cyanide has been investigated by Andres et al. [28, 29] at the HF level of theory using the 4-31G basis set. They determined the activation energy to be 333 kJ mol^{-1} and the energy of reaction to be 0.04 kJ mol^{-1} for the formation of the NH_3/HCN complex. A potential energy hypersurface was established. The first step was the isomerization of the *E* (*trans*) to *Z* (*cis*) isomer followed by a 1,3-shift of the hydrogen bonded to the imino nitrogen to the amino nitrogen accompanied by the cleavage of the bond between the carbon and the amino nitrogen. At present, no computational studies of water-mediated or solution phase decomposition of formamidine to yield hydrogen cyanide and ammonia have been reported. In this work, a detailed study of the unimolecular and the water-mediated (mono- and di-hydrated complexes) decomposition of formamidine, both in the gas phase and in solution, is presented. Because of the size of the system, it is possible to perform the calculations at high levels of theory, such as the Gaussian-n theories, which are known to give reliable energetics. The role played by water as a catalyst as well as its role as solvent on the

energetics and mechanism of this reaction were also investigated. The results serve as a benchmark for similar reactions in solution for more complex systems, few of which have been studied in depth computationally.

3.2 Computational Method

The MUNgauss [30] computational program was used for most of the geometry optimizations at HF/6-31G(d) for reactants, products, and transition state structures. All other calculations were performed with either Gaussian 98 [31] or Gaussian 03 [32]. The geometries of all reactants, transition states, and products were fully optimized at the restricted HF and second-order Moller-Plesset (MP2) levels of theory using the 6-31G(d) and 6-31+G(d) basis sets. Energies have also been calculated using the Gaussian-n theories, G1, G2, G2MP2, G3, G3B3, G3MP2, and G3MP2B3. B3LYP/6-31+G(d) density functional theory calculations were also performed. Intrinsic Reaction Coordinate (IRC) analysis was carried out for all transition states. Free energies of reaction and activation in aqueous solution for the unimolecular decomposition reaction of formamidine were calculated with the PCM (polarizable continuum model) as implemented in Gaussian03. The choice of cavity in PCM is important because the computed energies and properties depend on the cavity size. By default, the PCM model builds up the cavity using the United Atom (UA) model, i.e., by putting a sphere around each solute heavy atom; hydrogen atoms are enclosed in the sphere of the atom to which they are bonded. The three UA models which are available in Gaussian03 are not suitable for reactions involving

hydrogen transfers and hence were not used. In this study, the PCM solvent calculations were performed using the KLAMT cavity. All free energy calculations involving solvation were performed using both gas and solution phase structures optimized at the RHF/6-31G(d) and MP2/6-31G(d) levels. Other cavities were also investigated (UFF and PAULING) but it was impossible to obtain all the optimized structures for these models. Single point calculations based on gas phase optimized geometries using UFF and PAULING cavities gave results which were very close to the KLAMT cavity.

3. 3 Results and Discussion

The results for the decomposition reaction of formamidine and its mono- and di-hydrated complexes in the gas phase and in aqueous solution are given in Tables 3.1 to 3.3.

3. 3. 1 Thermodynamic Results for the Decomposition of Formamidine.

The thermodynamic properties for the decomposition reaction of formamidine are listed in Table 3.1. The reaction for the decomposition of formamidine is found to be endothermic using Gaussian-n theories (with enthalpies of 26.6 to 30.1 kJ mol⁻¹) and less endothermic for the formation of the NH₃/HCN complex (4.8 to 9.3 kJ mol⁻¹). These results do not differ by more than 4.5 kJ mol⁻¹, which is within the expected error of the Gaussian-n theories. The HF/6-31G(d) enthalpies are consistently in better agreement with the Gaussian-n theories than

the MP2/6-31G(d) enthalpies. The difference between the enthalpy for the formation of the complex and the formation of separated products ($-22.5 \text{ kJ mol}^{-1}$ at G3) represents the stability of the NH_3/HCN complex. Adding a single water molecule has little effect on the enthalpy of reaction (6.5 vs 6.7 kJ mol^{-1} at G3). Addition of the second water molecule has a larger effect on the enthalpy of reaction which is now only slightly endothermic (0.6 kJ mol^{-1} at G3). The decomposition reaction of formamidine with separated products is more exergonic than the formation of the HCN/NH_3 complex. This difference is simply due to the increase in the entropy for the formation of the separated products. Except for the HF solution results, all other levels of theory predict the reaction to be exergonic, where formation of the separated species ($\text{HCN} + \text{NH}_3$) is favoured over formation of the complex (HCN/NH_3).

3.3.2 Activation energy, Enthalpy and Free Energies of Activation, for the Decomposition of Formamidine.

3.3.2.1 The Gas Phase Unimolecular Decomposition of Formamidine

The geometries for the reactant, transition state, and products for the unimolecular gas phase decomposition of formamidine are shown in Scheme 3.1 along with the atom numbering scheme. Vibrational analysis for the two structures in Figure 3.1, shows that the C_s structure, used in a previous study [28], has one imaginary frequency, and the nearly planar C_1 structure to be a minimum. From these results, the equilibrium structure with C_1 symmetry was

considered as the starting point for this reaction in this study.

The IRC analysis of TS1 for the gas phase decomposition reaction of formamidine, is shown in Figure 3.2. The formamidine decomposition transition state can be described in terms of two bond ruptures (N_3-H_5 and N_4-C_1) and one bond formation (N_4-H_5). The transition state structure is highly strained, with an $N_3-C_1-N_4$ angle of 106.7° compared to 128.7° for the reactant. The product is a hydrogen bonded complex of ammonia with hydrogen cyanide. The activation energy for the unimolecular gas phase decomposition reaction of formamidine at various levels of theory are listed in Table 3.2. An activation energy of $333.1 \text{ kJ mol}^{-1}$ was reported for this reaction at HF/4-31G [28]. In this study, the activation energy is 331.5, 253.5, and $259.1 \text{ kJ mol}^{-1}$ at HF/6-31G(d), G1, and G3 levels, respectively. The free energy of activation for the reaction is 315.3, 253.8, and $259.4 \text{ kJ mol}^{-1}$ at HF/6-31G(d), G1, and G3 levels, respectively. These values are high and explain why rather extreme conditions are necessary to achieve this reaction experimentally.

Applying the solvent model yields a different mechanism for this reaction from the one in the gas-phase. The decomposition reaction in solution involves a two-step mechanism for both the unimolecular and the water-catalyzed decomposition. In solution, the first step involves transfer of a hydrogen from the imino nitrogen to the amino nitrogen forming a zwitterionic intermediate which is stabilized in solution. The transition state (TSa) and the zwitterionic intermediate for this reaction are shown in scheme 3.2.

The structures of the reactant and products for the unimolecular decomposition in the gas and solution phases are only slightly different. The first transition state structure for the reaction in solution is similar to the gas phase structure, but more product-like. For example the N_4-C_1 bond length increased from 1.47 Å in the gas phase to 1.53 Å in solution. The zwitterionic intermediate has a slightly longer N_4-C_1 bond length (1.58 Å) which increases significantly to 1.74 Å in the second transition state, giving a very product-like structure, with a very small activation energy (2.5 kJ mol⁻¹). In solution, the $N_3-C_1-H_2$ angle of the second transition state (**TSb**) is 140.3°, compared to 132.6° for the first transition state (**TSa**). The solvent models used in this study predict that the free energy of activation for the unimolecular decomposition would be reduced in aqueous solution. Using the PCM-KLAMT model, ΔG^\ddagger is lowered by 16 kJ mol⁻¹ and 18 kJ mol⁻¹ at HF/6-31G(d) and MP2/6-31G(d) levels, respectively. As given in Table 3.2, the free energy of activation for the unimolecular decomposition in solution is 237.9 kJ mol⁻¹ at MP2/6-31G(d) level of theory, not that different from the single point value of 236.1 kJ mol⁻¹.

3. 3. 2. 2 The Decomposition of Formamidine with one Water Molecule

Scheme 3.3 shows the one-step decomposition mechanism for the monohydrated complex of formamidine. This Scheme shows that the water molecule catalyses the decomposition by forming a nearly planar six-membered ring with formamidine. In this case, the $N_3-C_1-N_4$ angle in the transition state **TS2** shown in Figure 3.3 is 119.4° compared to 106.7° in the unimolecular reaction transition state (**TS1**, Scheme 3.1), an increase of 12.7°. The

activation energy, the enthalpy of activation and free energy of activation for this reaction are listed in Table 3.2. Adding one water molecule significantly relaxes the transition state, reducing the angle strain and consequently reduces the barrier height from 259.1 kJ mol⁻¹ to 169.4 kJ mol⁻¹ at the G3 level of theory. The free energy of activation for this system is 246.3, 178.3, and 182.4 kJ mol⁻¹ at HF/6-31G(d), G1, and G3 levels, respectively. These results clearly show that the kinetically difficult reaction takes place more easily when catalyzed by a single water molecule. The IRC analysis of the transition state **TS2** for this system is shown in Figure 3.4.

In solution, the one water-catalyzed decomposition reaction of formamidine also involves a two-step mechanism forming a zwitterionic intermediate as shown in Scheme 3.4. The structure of the reactant and products for the one water-catalyzed decomposition in both gas and solution phases also only differ slightly. As for the unimolecular reaction, the N₄-C₁ bond length increases from 1.50 Å in the gas phase to 1.55 Å in the solution phase in the first step of the reaction. In solution, the N₃-C₁-H₂ angle of the second transition state (**TSd**) is 141.4°, compared to 128.6° for the first transition state (**TSc**).

The zwitterionic intermediate has a slightly longer N₄-C₁ bond length (1.59 Å) which increases significantly to 1.77 Å in the second transition state, giving an even more product-like structure, with a very small activation energy (6.0 kJ mol⁻¹). In solution, the free energy of activation would further be reduced by 21 to 42 kJ mol⁻¹ (depending on the level of theory used). Free energy of activation for the single water-mediated decomposition in solution is

148.9 kJ mol⁻¹ at the MP2/6-31G(d) level, not significantly different from the single point value (157.6 kJ mol⁻¹) although the mechanism changes to a two step mechanism in solution.

3. 3. 2. 3 The Decomposition Reaction of Formamidinium with Two Water Molecules

Previous research on 1,3-hydrogen shift tautomerization reactions [20], has identified transition state structures in which two water molecules participate in the hydrogen transfer by forming an eight-membered ring. Addition of the second water molecule further relieves angle strain, lowering the activation energy for the reaction. However, all attempts to optimize the cyclic eight-membered dihydrated formamidinium transition state structure failed. This investigation did however lead to four different transition state structures containing a 6-membered ring as shown in Figure 3.5. Scheme 3.5 shows the only transition state (**TS3**) which connects the reactants and the products, as determined from the IRC analysis. This transition state is very similar in structure to **TS2** for the one water-mediated decomposition reaction of formamidinium. In **TS3**, one water molecule directly participates in the hydrogen transfer, as shown in Scheme 3.5. The second water molecule does not directly participate in hydrogen transfer but further stabilizes the transition state by hydrogen bonding to both an N₄ hydrogen and O₁₀.

Activation energies, enthalpies of activation and free energies of activation are listed in Table 3.2 for the decomposition of dihydrated formamidinium. Adding the first water molecule

reduces the activation energy by 89.7 kJ mol^{-1} at the G3 level. Adding the second water molecule reduces the activation energy of the formamidine decomposition to a much smaller extent, further decreasing the activation energy by only 18.3 kJ mol^{-1} . This is because the first water molecule catalyses the reaction whereas the second water molecule simply stabilizes the transition state by “solvation”. Thus, the water can both stabilize the transition state and act as a catalyst for these reactions. The present results support a single water-mediated decomposition mechanism and suggest that a two water-mediated transition state does not appear to exist, in the gas phase or in aqueous solution, for this reaction.

3.3.3 Intermolecular Hydrogen-Transfer Transition States

Three additional transition states were found, **TS4a**, **TS4b**, and **TS5**, which represent intermolecular hydrogen-transfer between formamidine and either one or two water molecules and do not result in product formation. Activation energies, enthalpies of activation and free energies of activation for these intermolecular hydrogen-transfers are listed in Table 3.3. **TS4a** and **TS4b** are rotational isomers, differing from each other for the most part, with respect to the orientation of H_{12} (Figure 3.5). These two transition states involve the exchange of a hydrogen between formamidine and the two hydrogen-bonded water molecules through a six-membered cyclic structure. The IRC confirms that these transition states lead to the reactant in both directions. **TS5** has the highest activation energy among the transition states, and involves a hydrogen (H_5) transfer from a single water to N_4 of formamidine followed by rotation

about the C-N bond and transfer of a hydrogen (H_g) back to water (Figure 3.5). The IRC confirms that this transition state goes to the reactant in both directions.

3.3.4 Intrinsic Reaction Coordinate (IRC) Analysis

IRC analysis was carried out for all transition states. The IRC analysis for the gas phase transition state, **TS1** (Figure 3.2) and the monohydrated transition state, **TS2** (Figure 3.4), confirmed that these transition states connected the expected reactants and products. However, the IRC analysis of the dihydrated transition states, **TS3**, **TS4** and **TS5**, using the CalcFC option, did not give the complete reaction pathway in several cases, stopping prematurely as shown in Figures (3.6-3.8). Initially, this suggested the existence of an intermediate. Further analysis was carried out using the CalcFC or CalcALL options with structures optimized at different levels of theory and basis set. None of these options consistently gave a complete reaction pathway. In each of the cases in which the complete reaction pathway was obtained, shoulders are found to exist, as can be seen in Figures (3.6-3.8). Such shoulders for some of the options listed above may be indicative of zwitterionic species, unstable in the gas phase, but stable in solution phase.

3.4 Conclusions

The high degree of stability of formamidine toward decomposition to ammonia and hydrogen cyanide is confirmed by this investigation. We have performed *ab initio* calculations on water-mediated hydrogen transfer in the decomposition reaction of formamidine. Optimized geometries for reactants, transition states, and products were determined at the HF/6-31G(d) and MP2/6-31G(d) levels of theory. Energies were also determined at the G1, G2, G2MP2, G3, G3B3, G3MP2, and G3MP2B3 levels of theory. The role of water in the decomposition reaction of formamidine was examined. Intrinsic Reaction Coordinate IRC analysis was carried out for all transition state structures to obtain the complete reaction pathway. Activation energies, enthalpies and free energies of activation were also calculated for each reaction pathway. The G3 level of theory predicts the gas phase decomposition of formamidine to have a high activation energy of 259.1 kJ mol⁻¹. Adding one water molecule catalyses the reaction by forming a cyclic hydrogen bonded transition state, reducing the barrier to 169.4 kJ mol⁻¹ at G3 level. Addition of a second water, which acts as a “solvent” molecule, further reduces the barrier to 151.1 kJ mol⁻¹ at G3 level. A two water-mediated transition state has not been found despite a thorough investigation. These values are still high and explain why rather extreme conditions are necessary to achieve this reaction experimentally.

Thermodynamic properties (ΔE , ΔH , and ΔG) for each reaction pathway studied were also calculated. The G3 heats of reaction (ΔE) of the gas phase decomposition of formamidine, its complex with one water molecule and its complex with two water molecules are 0.9, 2.2,

and -5.1 kJ mol^{-1} , respectively. The G3 heat of reaction for the gas phase decomposition to yield separated products is 22.3 kJ mol^{-1} . Free energies of reaction and of activation in aqueous solution were calculated with PCM using the KLAMT cavity model. At MP2 the formamidine reaction is found to be exergonic in aqueous solution and to favour formation of the separated products ($\text{NH}_3 + \text{HCN}$).

The decomposition reaction of formamidine in solution involves a two-step mechanism for both the unimolecular and the water-catalyzed decomposition. The solvent model predicts a significant lowering of the free energy of activation, $16\text{-}18 \text{ kJ mol}^{-1}$ for the unimolecular reaction and $21\text{-}42 \text{ kJ mol}^{-1}$ for the water-mediated reaction in aqueous solution relative to the gas phase. The activation energies and the heats of reaction calculated at the Gaussian-n theories agree to within 10 kJ mol^{-1} .

3.5 References

- (1) Grout, R. J. J. *The Chemistry of Amidines and Imidates*. Edited by Patai, S. John Wiley & Son, New York, **1975**.
- (2) Yamabe, T.; Yamashita, K.; Kaminoyama, M.; Koizumi, M.; Tachibana, A.; *J. Phys. Chem.* **1984**, *88*, 1459.
- (3) Scheiner, S.; Kern, C. W.; *J. Am. Chem. Soc.* **1979**, *101*, 15, 4081.
- (4) Hroudá, V.; Florian, J.; Polašek, M.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 4742.
- (5) Hroudá, V.; Florian, J.; Hobza, P. *J. Phys. Chem.* **1993**, *97*, 1542.
- (6) Nguyen, K. A.; Gordon, M. S.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 1596.
- (7) Zielinski, T. J.; Poirier, R. A. *J. Comp. Chem.* **1984**, *5*, 5, 466.
- (8) Poirier, R. A.; Majlessi, D.; Zielinski, T. J. *J. Comp. Chem.* **1986**, *7*, 4, 464.
- (9) Poirier, R. A.; Yu, D.; P. Surjan, R.; *Can. J. Chem.* **1991**, *69*, 1589.
- (10) Kim, Y.; Lim, S.; Kim, H.-H.; Kim, Y. *J. Phys. Chem. A* **1999**, *103*, 617.
- (11) Pecul, M.; Leszczynski, J.; Sadlej, J. *J. Phys. Chem. A* **2000**, *104*, 8105.
- (12) Wiberg, K. B.; Rablen, P. R. *J. Am. Chem. Soc.* **1995**, *117*, 2201.
- (13) Czernek, J. *J. Phys. Chem. A* **2003**, *107*, 3952.
- (14) Wang, X.-C.; Nichols, J.; Feyereisen, M.; Gutowski, M.; Boatz, J.; Haymet, A. D. J.; Simons, J. *J. Phys. Chem.* **1991**, *95*, 10419.
- (15) Dziekonski, P.; Sokalski, W. A.; Podolyan, Y.; Leszczynski, J. *Chem. Phys. Lett.* **2003**,

367, 367.

- (16) Podolyan, Y.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2002**, 106, 12103.
- (17) Zhanpeisov, N. U.; Leszczynski, J. *J. Phys. Chem. A* **1999**, 103, 8317.
- (18) Shimoni, L.; Glusker, J. P.; Bock, C. W. *J. Phys. Chem.* **1996**, 100, 2957.
- (19) Kim, Y.; Lim, S.; Kim, Y. *J. Phys. Chem. A* **1999**, 103, 6632.
- (20) Zhang, Q; Bell, R.; Truong, T. N. *J. Phys. Chem.* **1995**, 99, 592.
- (21) Lim, J-H.; Lee E. K.; Kim, Y. *J. Phys. Chem. A* **1997**, 101, 2233.
- (22) Nagaoka, M.; Okuno, Y.; Yamabe, T. *J. Am. Chem. Soc.* **1991**, 113, 769.
- (23) Tortajada, J.; Leon, E.; Luna, A.; Mo, O.; Yanez, M. *J. Phys. Chem.* **1994**, 98, 12919.
- (24) Kim, Y.; *J. Phys. Chem. A* **1998**, 102, 3025.
- (25) Bell, R. L.; Truong, T. N. *J. Phys. Chem. A* **1997**, 101, 7802.
- (26) Flinn, C.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2003**, 107, 11171.
- (27) Frederico, L. A.; Kundel, T. A; Shaw, B. R. *Biochemistry* **1990**, 29, 2532.
- (28) Andres, J.; Krechl, J.; Carda, M.; Silla, E. *In. J. Quantum. Chem.* **1991**, XL, 127.
- (29) Andres, J.; Beltran, A; Carda, M.; Krechl, J.; Monterde, J.; Silla, E. *J. Mol. Struct.: THEOCHEM* **1992**, 254, 465.
- (30) Poirier, R. A.; *MUNgauss* (Fortran 90 version) [computer program]. Chemistry Department, Memorial University of Newfoundland, St. John's, Newfoundland, A1B 3X7 Canada. With contributions from Brooker, M. ; Bungay, S. D.; El-Sherbiny, A.; Gosse, T.; Keefe, D.; Pye, C. C.; Reid, D.; Shaw, M.; Wang, Y.; Xidos, J.; (b) Colonna, F.; Jolly, L-H.; Poirier, R. A.; Angyan, J.; Jansen, G.; *Comput. Phys. Commun.* **1994**, 81, 293-317.

(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman J. R. ; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M. ; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A. ; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M. ; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C. ; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; *Gaussian 98*, Revision A.11.3, Gaussian, Inc., Pittsburgh PA, **2002**.

(32) Frisch,, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R. ; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H. ; Hada, M.; Ehara, M.; Toyota, K.;Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y. ; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C. ; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J. ; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G. ; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.;

Piskorz, P.; Komaromi, I. ; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.;
Nanayakkara, A. ; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.;
Gonzalez, C.; Pople, J. A.; *Gaussian 03*, Revision B.05, Gaussian, Inc., Pittsburgh PA, **2003**.

TABLE 3.1: Thermodynamic properties for the decomposition reaction of formamidine (in kJ mol⁻¹) at 298.15 K.

	Formamidine ¹			Formamidine ²			Formamidine/H ₂ O ²			Formamidine/2H ₂ O ²		
	ΔE	ΔH	ΔG^a	ΔE	ΔH	ΔG^b	ΔE	ΔH	ΔG^c	ΔE	ΔH	ΔG^d
HF/6-31G(d)	36.9	23.5	-20.2	5.0	-2.5	-15.6	6.9	-1.2	-10.5	1.6	-7.7	-21.3
Solution ^e			-18.5			4.8						
			-25.1			3.1						
MP2/6-31G(d)	26.9	13.5	-30.5	-11.4	-18.2	-30.3	-9.9	-17.6	-24.9	-5.5	-14.8	-30.2
Solution ^e			-23.2			-20.5						
			-35.5			-23.3						
G1	22.1	28.8	-14.3	-0.8	4.8	-9.0	-1.1	3.5	-6.5	-8.5	-2.7	-17.0
G2	23.2	29.8	-13.3	2.6	8.2	-5.6	3.3	7.9	-2.1	-2.8	2.9	-11.4
G2MP2	23.5	30.1	-13.0	3.7	9.3	-4.5	4.3	8.9	-1.1	-1.9	3.9	-10.4
G3	22.3	29.0	-14.1	0.9	6.5	-7.3	2.2	6.7	-3.2	-5.1	0.6	-13.7
G3B3	21.0	27.8	-15.6	1.7	6.6	-5.0	3.5	7.0	0.0	-5.1	0.0	-11.5
G3MP2	20.8	27.5	-15.6	0.4	6.0	-7.8	1.7	6.2	-3.7	-5.2	0.6	-13.7
G3MP2B3	19.8	26.6	-16.9	1.6	6.5	-5.1	3.4	6.9	-0.1	-4.9	0.2	-11.3

¹ For the formation of separated products. ² For the formation of the HCN/NH₃/nH₂O complex (n = 0-2).

^a ΔS values range from 0.145 to 0.148 kJ mol⁻¹ K⁻¹. ^b ΔS values range from 0.039 to 0.046 kJ mol⁻¹ K⁻¹.

^c ΔS values range from 0.017 to 0.033 kJ mol⁻¹ K⁻¹. ^d ΔS values range from 0.035 to 0.052 kJ mol⁻¹ K⁻¹. ^e Values are reported in the order;

PCM-KLAMT model for single point calculation and for optimized structures. In all case $\Delta G = \Delta \Delta G(\text{thermal correction}) + \Delta G_{\text{solv}}$.

TABLE 3.2: Activation energies, enthalpies of activation and free energies of activation for the decomposition reaction of formamidine (in kJ mol⁻¹) at 298.15 K.

	Formamidine			Formamidine /H ₂ O			Formamidine/2H ₂ O		
	ΔE_{act}	ΔH^\ddagger	ΔG^\ddagger ^a	ΔE_{act}	ΔH^\ddagger	ΔG^\ddagger ^b	ΔE_{act}	ΔH^\ddagger	ΔG^\ddagger ^c
HF/6-31G(d)	331.5	314.3	315.3	244.0	226.7	246.4	217.4	203.3	219.7
Solution ^d			296.1			212.5			
			299.0			204.6			
MP2/6-31G(d)	274.1	255.4	255.7	173.1	152.8	169.9	156.2	139.6	152.6
Solution ^d			236.1			157.6			
			237.9			148.9			
G1	253.5	252.7	253.8	165.3	158.0	178.3	146.9	141.0	158.0
G2	257.6	256.8	257.9	168.8	161.5	181.9	151.4	145.5	162.5
G2MP2	257.2	256.3	257.4	168.8	161.5	181.9	151.6	145.6	162.7
G3	259.1	258.2	259.4	169.4	162.1	182.4	151.1	145.2	162.2
G3B3	256.6	256.1	256.6	164.9	158.6	174.7	145.5	140.9	154.5
G3MP2	257.9	257.1	258.2	172.7	165.4	185.8	154.7	148.8	165.8
G3MP2B3	255.4	254.9	255.4	168.1	161.8	177.8	148.7	144.0	157.7

^a ΔS^\ddagger values range from -0.001 to -0.004 kJ mol⁻¹ K⁻¹. ^b ΔS^\ddagger values range from -0.041 to -0.068 kJ mol⁻¹ K⁻¹. ^c ΔS^\ddagger values range from -0.04 to -0.06

kJ mol⁻¹ K⁻¹. ^dValues are reported in the order; PCM-KLAMT model for single point calculation and for optimized structures. In all case

$$\Delta G = \Delta \Delta G(\text{thermal correction}) + \Delta G_{solv}.$$

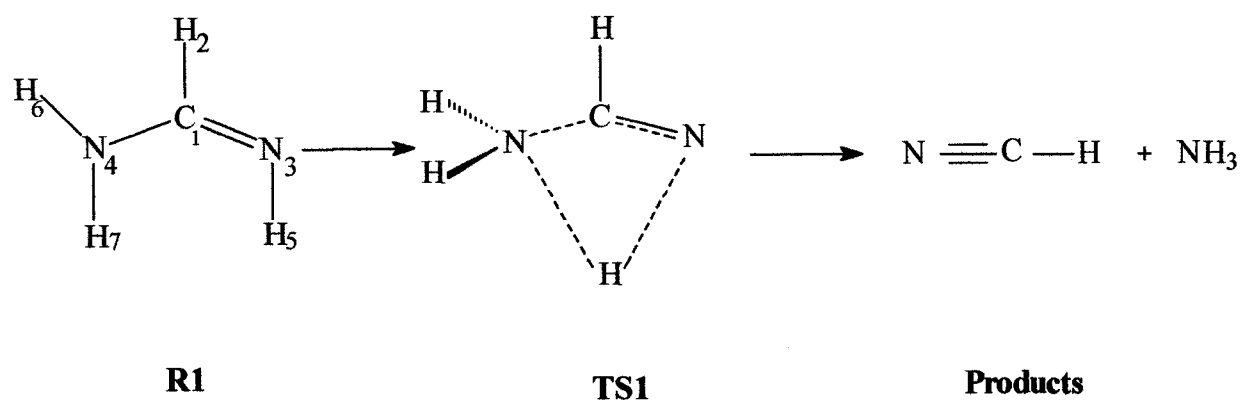
TABLE 3.3: Activation energies, enthalpies of activation and free energies of activation (in kJ mol⁻¹) at 298.15 K for the intermolecular hydrogen-transfers in formamidine/2H₂O (TS4a, TS4b, and TS5).

	TS4a			TS4b			TS5		
	ΔE_{act}	ΔH^\ddagger	ΔG^\ddagger	ΔE_{act}	ΔH^\ddagger	ΔG^\ddagger	ΔE_{act}	ΔH^\ddagger	ΔG^\ddagger
HF/6-31G(d)	211.3	195	215	206.6	190.2	210.6	236.5	233.1	246.7
MP2/6-31G(d)	142.1	121.5	137.3	135.6	114.7	130.8	156.2	139.6	152.6
G1	134.1	126.3	147.2	128.4	120.3	141.7	155.9	150.4	164.7
G2	136.8	129	149.9	132.3	124.2	145.6	160.4	154.9	169.2
G2MP2	136.7	128.9	149.8	132.3	124.3	145.6	160.6	155	169.3
G3	137	129.2	150.1	132.3	124.2	145.6	160.2	154.6	168.9
G3B3	129.6	123.4	138.4	124.2	117.9	133.2	145.5	140.9	154.5
G3MP2	142.5	134.7	155.6	138.1	130	151.4	163.7	158.2	172.5
G3MP2B3	134.9	128.7	143.7	129.8	123.4	138.7	148.7	144	157.7

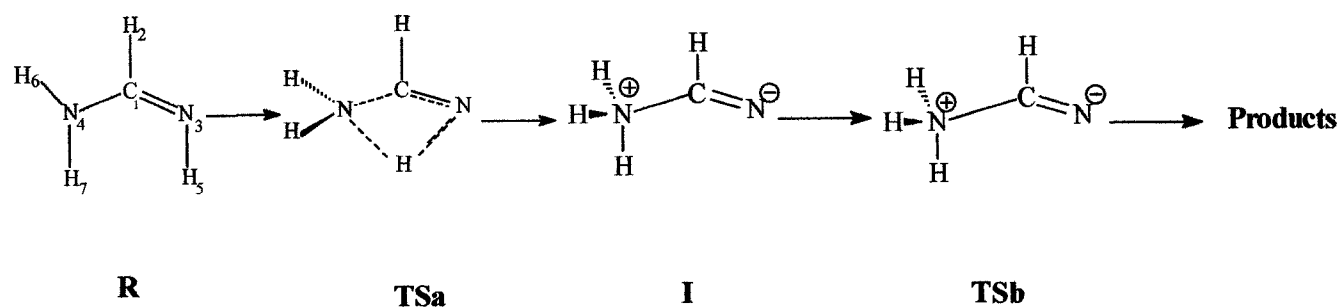
ΔS^\ddagger values range from -0.05 to -0.07 kJ mol⁻¹ K⁻¹ for TS4a.

ΔS^\ddagger values range from -0.05 to -0.07 kJ mol⁻¹ K⁻¹ for TS4b.

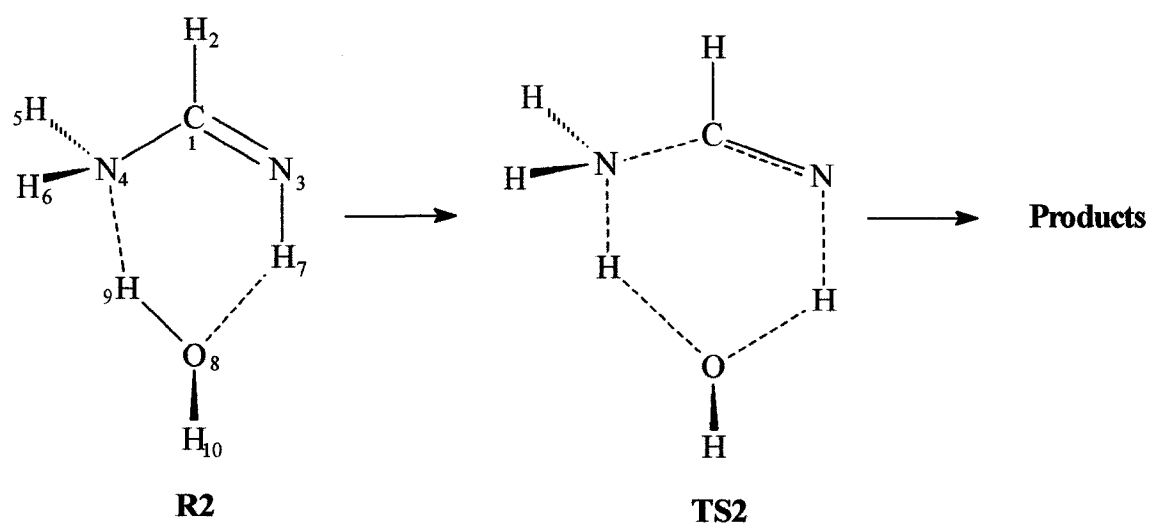
ΔS^\ddagger values range from -0.03 to -0.05 kJ mol⁻¹ K⁻¹ for TS5.



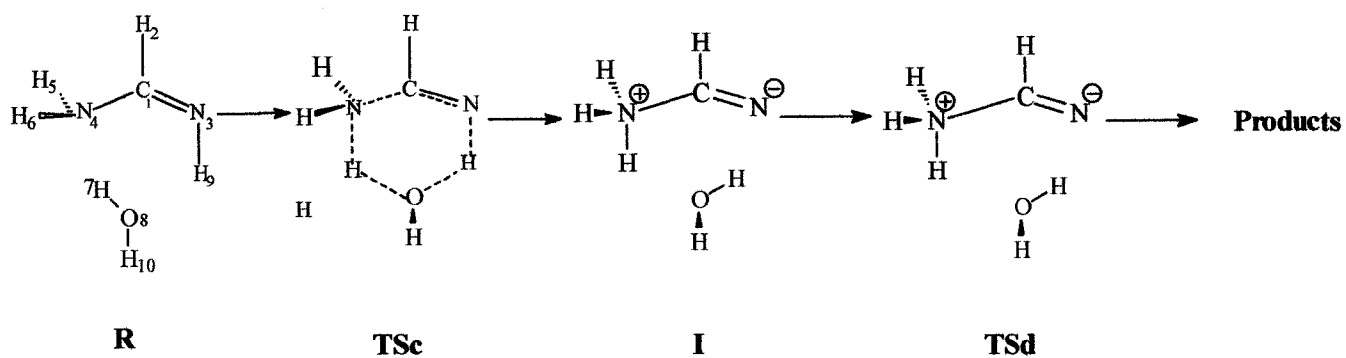
Scheme 3.1: The gas-phase unimolecular decomposition reaction of formamidine.



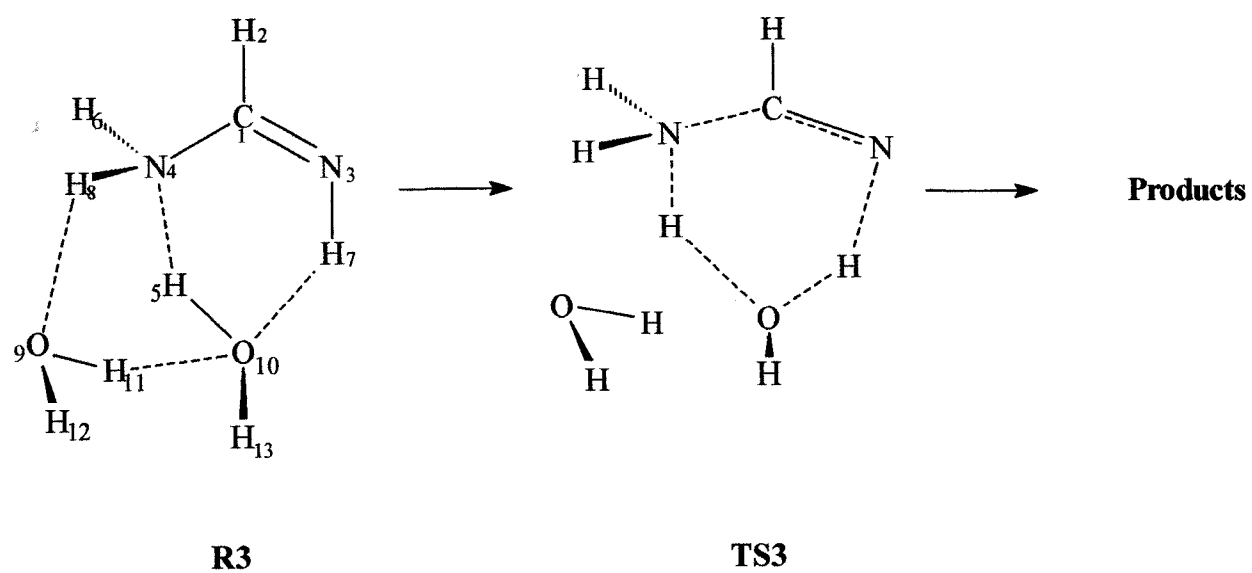
Scheme 3.2: The decomposition reaction of formamidine in solution.



Scheme 3.3: The decomposition reaction of formamidine with one water molecule.



Scheme 3.4: The decomposition reaction of formamidine with one water molecule in solution.



Scheme 3.5: The decomposition reaction of formamidine with two water molecules.

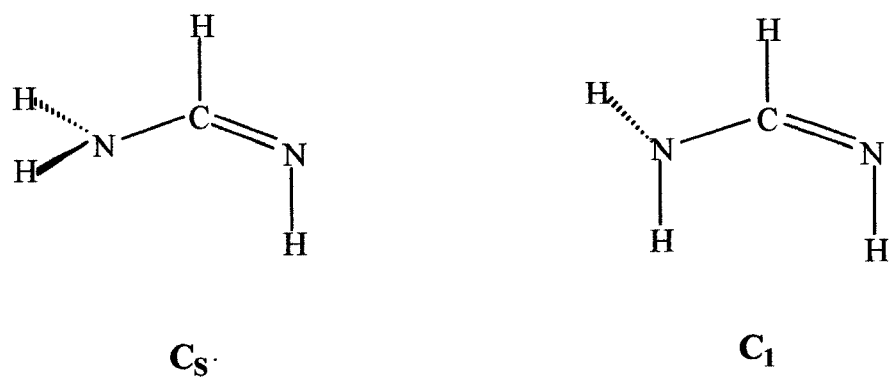


Figure 3.1: The C_s and C_1 structures of formamidinium.

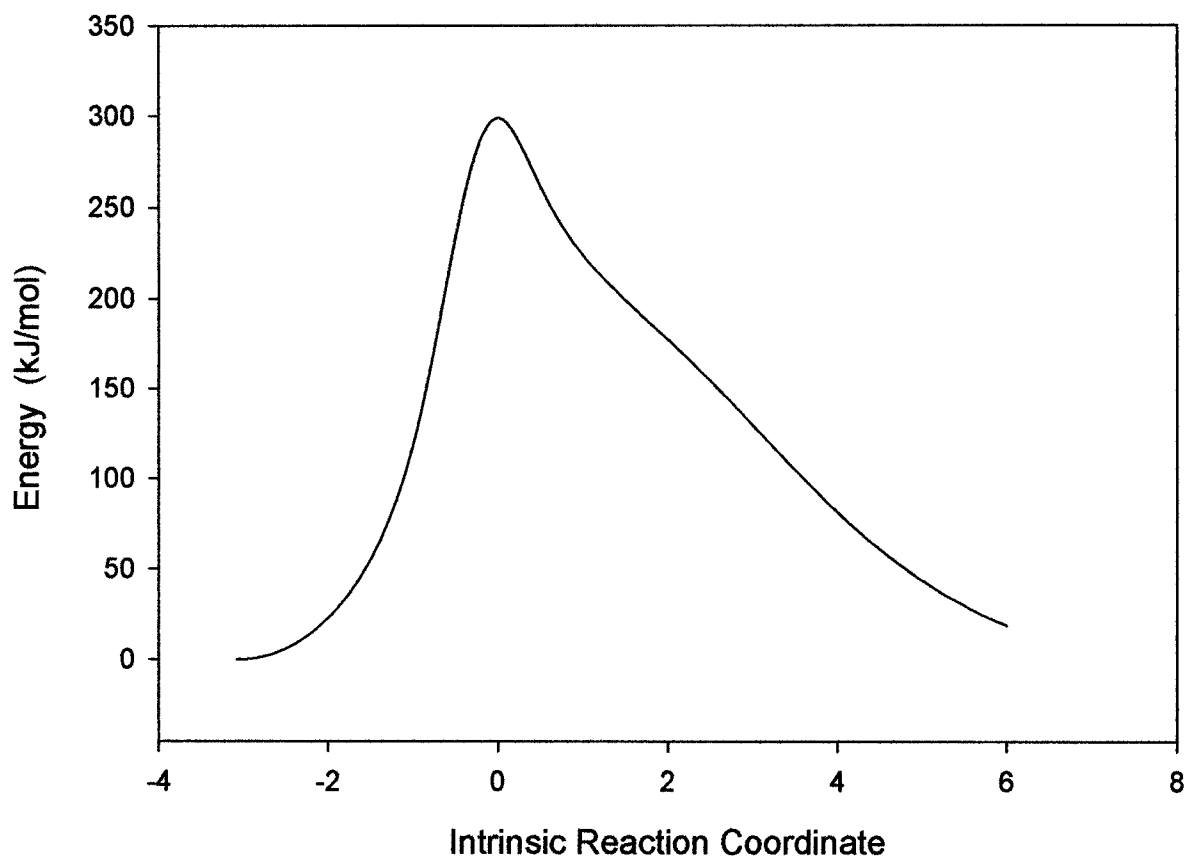


Figure 3.2: Energy vs the IRC for the gas phase decomposition of formamidine at HF/6-31G(d) with the CalcFC option. The energy of the reactant is taken to be zero.

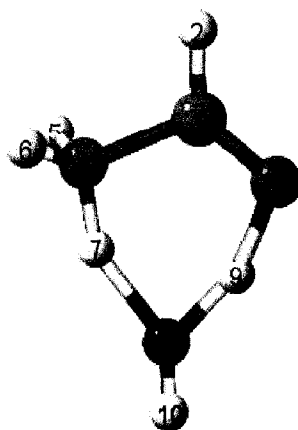


Figure 3.3: The transition state structure for the decomposition reaction of formamidine with one water molecule.

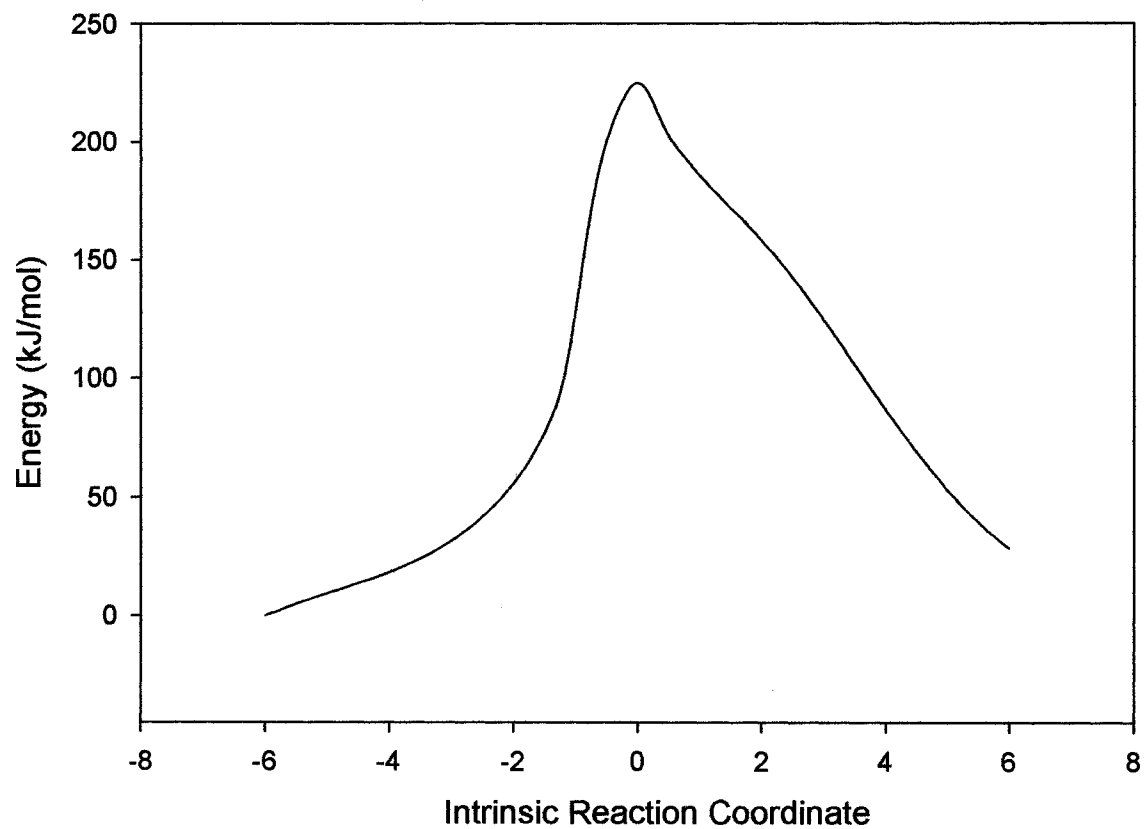
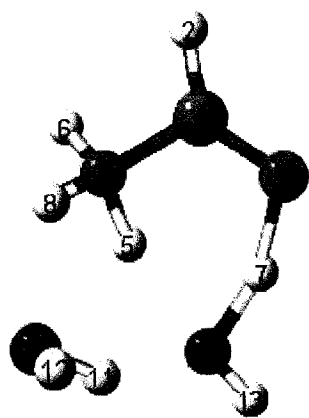
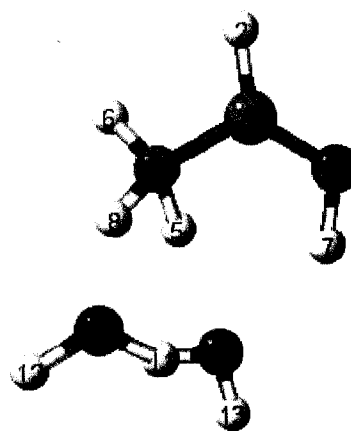


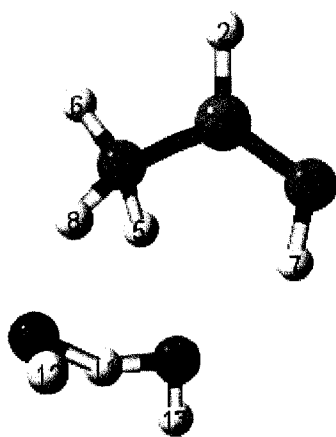
Figure 3.4: Energy vs the IRC for the decomposition of monohydrated formamidine at HF/6-31G(d) with the CalcFC option. The energy of the reactant is taken to be zero.



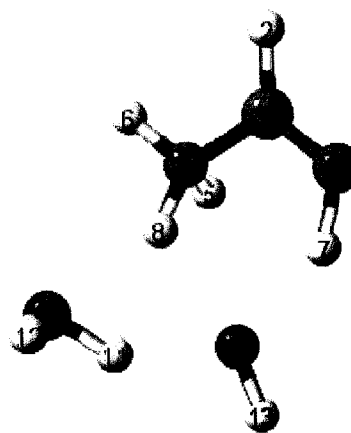
TS4a



TS3



TS4b



TS5

Figure 3.5: The transition state structures for dihydrated formamidine.

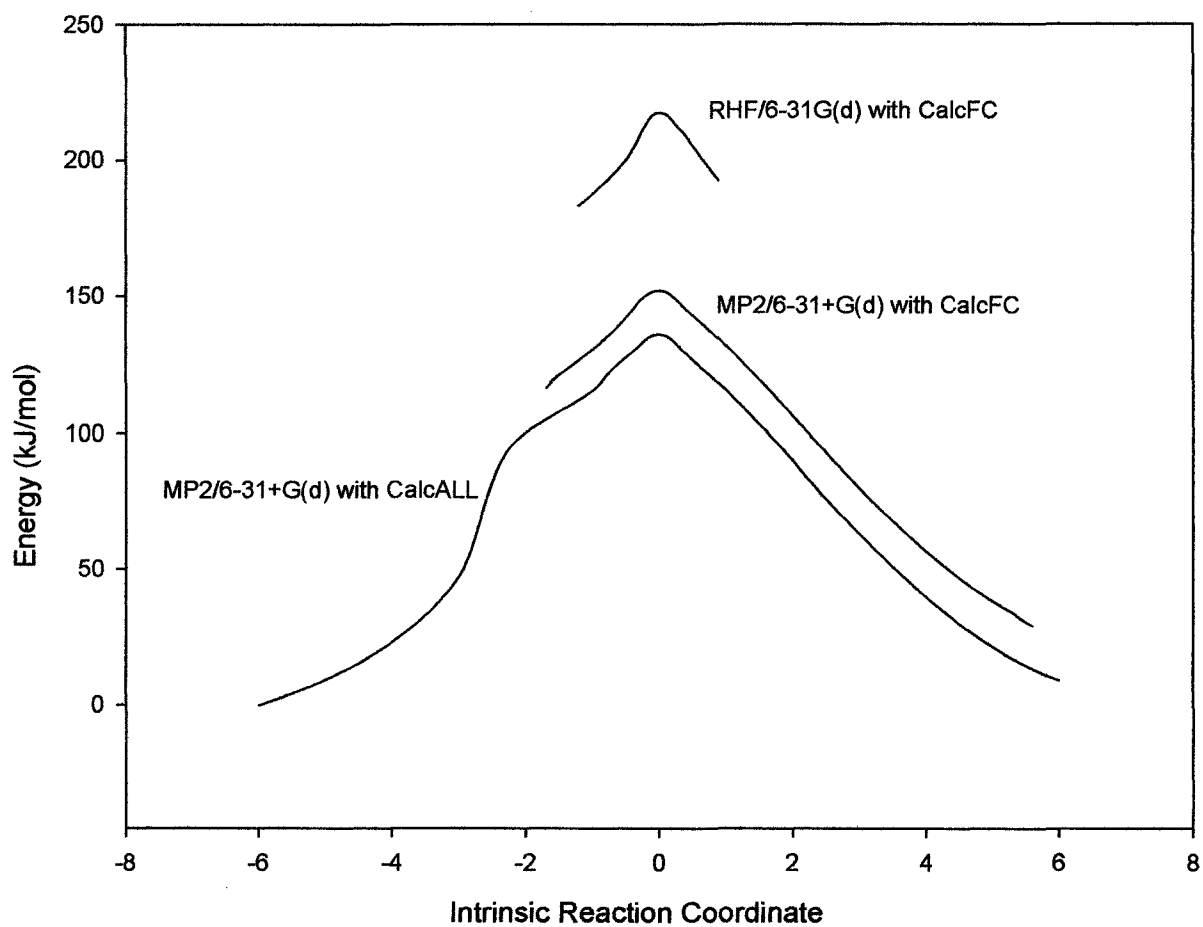


Figure 3.6: Energy vs the IRC for the decomposition of dihydrated formamidine, TS3. The energy of the reactant is taken to be zero.

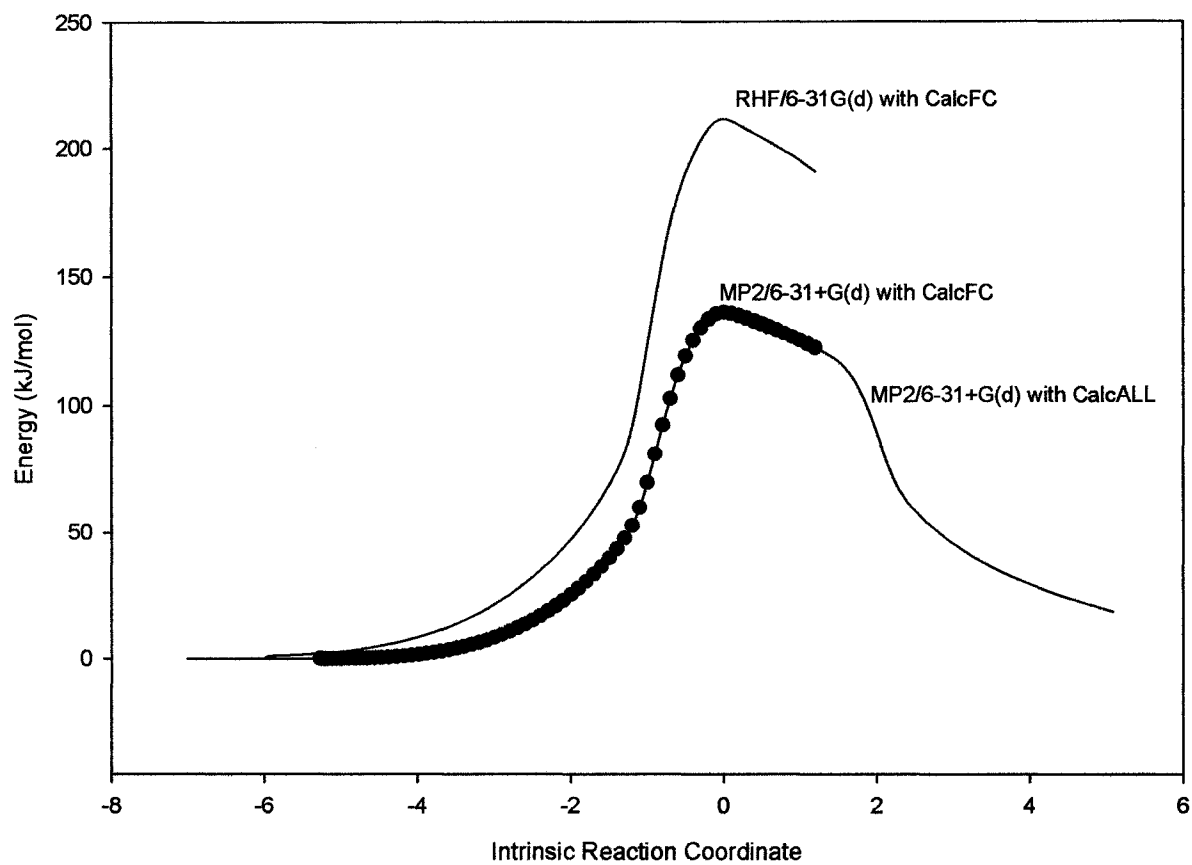


Figure 3.7: Energy vs the IRC for water/formamidinium hydrogen exchange, TS4. The energy of the reactant is taken to be zero.

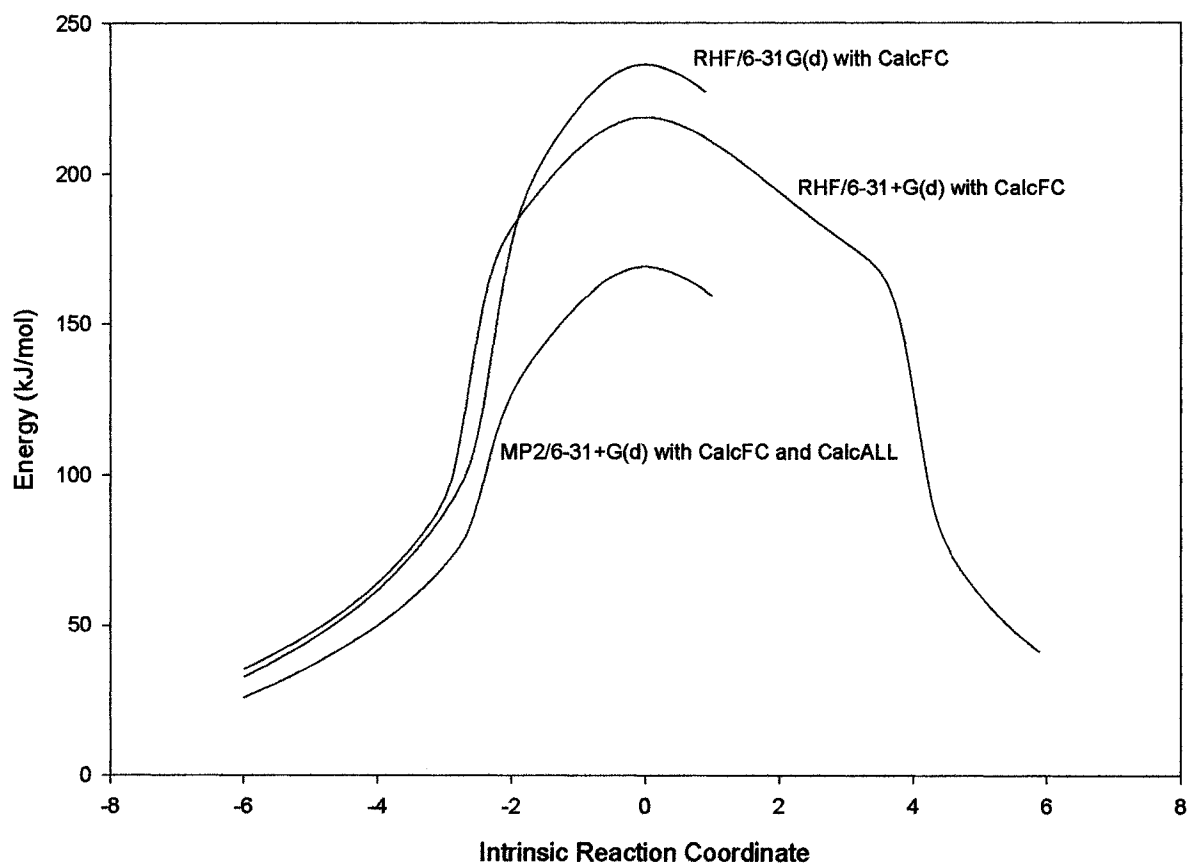


Figure 3.8: Energy vs the IRC for water/formamidinium hydrogen exchange, TS5. The energy of the reactant is taken to be zero.

CHAPTER 4

Computational Study of the Deamination Reaction of Cytosine with H₂O and OH⁻ ⁵

4.1 Introduction

Cytosine (Cyt), one of the pyrimidine bases, occurs naturally in many nucleic acids, DNA, and RNA and is the most unstable of the DNA bases. It is chemically bound to a sugar moiety and interacts with other nucleic acid bases via hydrogen bonds, most frequently with guanine [1]. Cytosine is also a parent compound of various modified nucleosides and nucleotides. Some of them occur naturally, and others are the product of chemical reactions of nucleic acids with various mutagenic agents [2]. Among nucleic acid bases, cytosine is the most alkaline in aqueous solution ($pK_a = 4.6$). This property plays an important role in many biochemical processes [3]. Cytosine, its nucleosides and nucleotides, and many of its derivatives

⁵ Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2006**, 110, 8227–8234.

have been extensively studied experimentally [4–18], as well as computationally, in both the gas phase [17–42] and the aqueous phase [43–49].

The nucleic acid bases have tremendous versatility in the formation of hydrogen bond complexes because of the presence of numerous hydrogen bond donor and acceptor groups. These interactions are responsible for maintaining the genetic code [3]. Tautomerism is fundamentally important to the structure and functioning of nucleic acid bases. Interest in this area is due to the fact that the formation of rare tautomers can induce alterations in the normal base pairing, leading to spontaneous mutations in the genetic code [50]. Hence, a great deal of research has been carried out on the tautomerism of nucleic acid bases using both experimental and theoretical approaches. Conversion of one tautomer to another is generally the result of proton transfer reactions whose activation energy barriers may control the formation of higher-energy tautomeric forms (rare tautomers).

Purine and pyrimidine bases exist in several different tautomeric forms which differ from each other mainly in the position of one of the hydrogens which may be bound to the exocyclic nitrogen or oxygen atom or one of the ring nitrogen atoms. It has been suggested that conversion between tautomers can occur as a result of a simultaneous double-proton transfer, but such conversions are energetically rather unfavorable [36]. Numerous computational studies have discussed the tautomerism of the cytosine molecule [18,34–38,45–49] which have provided a reliable picture of the relative stability of its tautomers, both in the gas phase and in

solution. Using infrared spectrometry, it has been found that, in Ar and N₂ matrices at 15 K, cytosine exists as a mixture of the “normal” amino-oxo (a-o), the canonical form found in DNA, and the “rare” amino-hydroxy (a-h) tautomer [6]. In addition to the canonical form and the (a-h), there are imine tautomers. Since the tautomers are very close in energy, their relative stabilities are very sensitive to the level of theory [35,36]. It appears that the canonical form is the local minimum at almost all theoretical levels [38]. DFT calculations favor slightly the amino-oxo form [35,51]. The a-h tautomer has been predicted to predominate in the gas phase using a coupled cluster technique [49]. A microwave investigation has identified three of the six possible tautomers [17].

The relative energies of the cytosine–water conformers have been computed at the MP2/6-31G(d) level of theory and the results have shown that the cytosine (keto)-water complex is more stable than the cytosine (enol)-water complex by 4.5 kJ mol⁻¹ [39]. It has also recently been shown that the amino-oxo tautomeric form of the cytosine-water complex is stable in both the ground and excited states [52]. By comparison with experimental data [6,17], it was shown that kinetic barriers in the gas phase are generally high (126–155 kJ mol⁻¹) which affects the tautomer population at T ≤ 500 K [47]. It must be mentioned that the interaction between the canonical form of cytosine and one and two water molecules has been investigated in several theoretical works [37–39,47,48]. Theoretical studies have shown that interaction with water molecules changes the relative energies of some or all tautomers, the canonical tautomer being better hydrated than the other tautomers [53]. Florian et al. [18] studied the relative

stabilities of tautomers of protonated cytosine in the gas phase and in a polar solvent at the *ab initio* HF, MP2, and polarizable continuum approximations. In addition, the infrared and Raman frequencies and intensities for cytosine and its two protonated forms were calculated.

Metal ions can also affect the relative stability of tautomers of the pyrimidine base cytosine. There have been several studies of the interaction of cytosine and its tautomers with metals [32,40–42]. Metalation can play a role in the formation of rare tautomers of nucleic acid bases and also affect the ability of the nucleobase to be protonated or deprotonated [42]. More recently, Prado et al. [54] studied the Mg^{2+} –cytosine complex interaction at RHF, MP2, DFT, and CCSD(T) levels of theory with the 6-31+G(d) basis set.

Interactions of other species such as hydrogen peroxide and hydroxyl radical have also been the object of recent studies. For example, authors have studied the interaction between the amino-oxo and *cis* amino-hydroxy tautomers of cytosine with one hydrogen peroxide molecule. The binding energies and harmonic vibrational frequencies for the complexes were calculated using the B3LYP/6-31++G(d,p) level of theory [21]. DFT calculations have shown that the addition of an OH radical to the C₅ and C₆ sites are thermodynamically and kinetically more favorable than addition to the N₃ and C₂ sites [20].

Spontaneous mutations (changes to the nucleotide sequences of DNA) can arise as a result of chemical changes to individual bases in DNA. One such chemical change is the

conversion of cytosine to uracil which is classified as a deamination reaction. In general, deamination refers to the loss of an amino group from a tetrahedral carbon with conversion to a carbonyl functional group. All the DNA bases, except thymine and uracil have amino groups, but only deamination of cytosine gives a base found in DNA and RNA. Hydrolytic deamination of cytosine yields uracil, as shown in Figure 4.1.

Uracil (Ura) is found in RNA and can only base pair with adenine [10]. In principle, six tautomeric forms of uracil are possible [35], and like cytosine, numerous studies on all tautomers of uracil have been reported [35,36,27,39,45,55]. If uracil is found in DNA, it poses a very serious problem. The cell, however, has a specific enzyme to remove it from DNA, called DNA uracil-*N*-glycosylase. The uracil formed by cytosine deamination is potentially mutagenic, changing the coding information during DNA replication and RNA transcription, resulting in altered base pairs in the genome [13]. Hydrolytic deamination is known to depend on pH and temperature [9–12]. It has been established that deamination of the DNA base cytosine is an extremely rare event under normal physiological conditions (40–100 deaminations in the human genome per day at pH 7.4), although the rate of deamination can be significantly increased in the presence of various reagents such as NO, HNO₂ and bisulfite. For example, bisulfite-induced deamination involving acid-catalyzed hydrolysis [14] has been shown to noticeably accelerate the rate of cytosine deamination. Other pathways that were studied include a diazotization pathway [56], and base-catalyzed [10], and acid-catalyzed [8] deaminations. Federico et al. [16] were able to determine the rate constant of cytosine deamination for single-

and double-stranded DNA under physiologically relevant conditions at 37 °C and pH 7.4 by a sensitive genetic assay. Their measured rate constants for single- and double-stranded DNA are 1×10^{-10} and $7 \times 10^{-13} \text{ s}^{-1}$, respectively, with an activation energy of $117 \pm 4 \text{ kJ mol}^{-1}$. This value agrees well with the value of 121 kJ mol^{-1} obtained over a 25 °C temperature range by Lindahl and Nyberg [57].

Duncan et al. [58] pointed out that C₅ methylation of cytosine increases the likelihood of spontaneous mutations. They showed that the rate of deamination of 5-methylcytosine (m⁵Cyt) is three to four times faster than that for cytosine. As a result, m⁵Cyt residues are considered to represent hotspots for spontaneous transition mutations [58]. The deamination of m⁵Cyt, which is involved in the regulation of gene expression, yields thymine which is naturally found in DNA. Replication of these deamination products will produce a C·G → T·A transition mutation [13,16].

Yao et al. [4] studied the hydrolysis of cytosine to uracil by yeast cytosine deaminase, a zinc metalloenzyme. In particular, they studied the catalysis of the deamination of the prodrug 5-fluorocytosine to form the anticancer drug 5-fluorouracil. Sponer et al. [5] studied the metal-mediated deamination of 1-methylcytosine and 1,5-dimethylcytosine with a cationic complex of Pt^{II} both experimentally and using DFT calculations. They also studied the deamination of cytosine with OH⁻ using the PCM model to account for solvation effects. This study represents the only computational study of the hydrolytic deamination reaction of cytosine which has been

reported [5]. However, their reported activation energy barrier ($213.4 \text{ kJ mol}^{-1}$) is not very close to the experimentally accepted value ($117 \pm 4 \text{ kJ mol}^{-1}$) [16,55].

This chapter represents a detailed computational study of the hydrolytic deamination of the DNA base cytosine with OH^- and H_2O and follows a similar study of the deamination of the related compound, formamidine (*E* isomer) with OH^- , H_2O , and H_3O^+ [59]. The deamination of the structurally related *E*(trans)-isomer of formamidine, as predicted, provided reaction pathways which proved to be very similar in terms of transition states, intermediates, and energetics to the reaction pathways of cytosine. As can be seen in Figure 4.2, the *E* isomer of formamidine forms part of the cytosine structure, notably in the region of cytosine where deamination of cytosine to uracil can take place.

4.2 Computational Method

The MUNgauss [60] computational program was used for most of the geometry optimizations at the HF/6-31G(d) level for reactants, products, and transition state structures. All other calculations were performed with Gaussian 03 [61]. The geometries of all reactants, transition states, intermediates, and products were fully optimized at the restricted HF, second-order Møller-Plesset (MP2) and B3LYP levels of theory using the 6-31G(d) basis set. Energies have also been calculated using the Gaussian-n theory, G3MP2. From our previous work [62], we found that the activation energies and the heats of reaction calculated using Gaussian-n theories (G1, G2, G2MP2, G3, G3MP2, G3B3, and G3MP2B3) all agreed to within 10 kJ mol⁻¹, which is within the reported error of the Gaussian-n theories in the literature. On the basis of these results, we chose the G3MP2 level of theory which is computationally the least expensive method to give reliable energetics. Single-point calculations at the B3LYP/6-31+G(d) and MP2/GTMP2Large levels of theory are also reported. All structures have been vibrationally characterized, i.e., checked for the absence of imaginary frequencies in the minima and for the presence of only one imaginary frequency in the transition states. The complete reaction pathway for all the mechanisms discussed in this paper have been verified using intrinsic reaction coordinate (IRC) analysis of all transition states. For each transition state, we optimized the structures at the last IRC points to positively identify the reactant and product to which each is connected.

4.3 Results and Discussion

Our previous results from the deamination reaction of formamidine [59] gave us a useful starting point for studying the mechanism of the deamination reaction of cytosine by providing good initial guesses for transition state structures that exist along the reaction pathways.

4.3.1 Deamination of Cytosine with H₂O in Pathways A and B

Computational studies predict that, in aqueous solution or the gas phase, cytosine protonation should occur at the N₃ position (see Figure 4.1) [18]. Computational studies have also shown that the interaction with water changes the relative energies of the other tautomers, the canonical tautomer being better hydrated than the other tautomers [53]. Deamination of cytosine with H₂O can follow one of two possible pathways designated as pathway A and pathway B. These pathways are closely related to those found for the deamination of formamidine [59]. The geometries for the reactant, intermediates, transition states, and product involved in pathway A, are shown in Figure 4.3. The relative energies of reactants, intermediates, transition states, and products for pathway A are shown in Figure 4.4.

Pathway A is a five-step mechanism. Initially, a hydrogen-bonded complex of cytosine with one water molecule forms the imine-oxo tautomer of cytosine (**IIa**) with an activation energy of 79.6 kJ mol⁻¹ at the G3MP2 level of theory. This is followed by nucleophilic attack by the water molecule on the C₄ atom and proton transfer from H₂O to the exocyclic imine

nitrogen (sp^2 nitrogen) of cytosine to form a tetrahedral intermediate (**I2a**). This is followed by two conformational changes of the initial intermediate to give two new intermediates (**I3a**, and **I4a**, in Figure 4.3), finally resulting in an intramolecular 1–3 proton transfer of the hydroxyl hydrogen to the amino group to yield a uracil–ammonia complex (**P1a**).

The geometries for the reactant, intermediates, transition states, and product involved in pathway B are shown in Figure 4.5. The relative energies of reactants, intermediates, transition states, and products for pathway B are shown in Figure 4.6. In pathway B, the water molecule attacks the C_4 atom with simultaneous proton transfer from H_2O to the amino group (sp^3 nitrogen) of cytosine producing the hydroxy-oxo tautomer of uracil and ammonia, **I1b**. A 1–3 proton shift of the hydroxy hydrogen to the sp^2 nitrogen of the tautomer forms the deamination product, the uracil–ammonia complex, as shown in Figure 4.5. Both complete reaction pathways have been verified using IRC analysis.

The mechanism for the deamination of cytosine with H_2O is very similar to that of the deamination of formamidine with H_2O . The most notable differences were for pathway A in which the initial tetrahedral intermediate undergoes several conformational changes and the initial intermediate is formed from a tautomer of cytosine in the first step of the reaction. Pathway B is very similar for both cytosine and formamidine.

The thermodynamic properties for the deamination of cytosine with H_2O are listed in Table 4.1. The deamination reaction of cytosine is found to be exothermic and exergonic at all

levels of theory for both separated species and for complex to complex, with negligible differences between them.

The activation energies, enthalpies of activation, and free energies of activation for the deamination of cytosine with H₂O at HF, MP2, and B3LYP levels of theory using the 6-31G(d) basis set and the G3MP2 level of theory for both pathways A and B are listed in Tables 4.2 and 3, respectively. The MP2 and B3LYP results are in reasonable agreement with G3MP2 results, differing by no more than 22 kJ mol⁻¹. The MP2/GTMP2Large results, for both the thermochemistry (Table 4.1) and barriers (Tables 4.2 and 4.3), are in excellent agreement with the G3MP2 values, differing by no more than 7 kJ mol⁻¹.

An activation energy of 117 ± 4 kJ mol⁻¹ was reported experimentally for this reaction [16,57]. In this study, the activation energies for the rate-determining step for pathways A and B are 221.3 and 260.3 kJ mol⁻¹, respectively, at the G3MP2 level of theory, respectively. These values are high compared with the experimental value and may explain why this reaction is an extremely rare event under neutral conditions. The activation energies for step three (TS3a) and step four (TS4a), which represent the formation of different conformers, are 1.7 and 6.6 kJ mol⁻¹ G3MP2, respectively. Pathway A is a more likely mechanism for the deamination with H₂O compared to pathway B due to a significantly lower activation energy (221.3 and 260.3 kJ mol⁻¹, respectively) for the rate-determining step. These activation energies are only slightly higher than those found for formamidine, 212.7 kJ mol⁻¹ for pathway A and 249.2 kJ mol⁻¹ for

pathway B (at the G2 level of theory) [59]. It can be seen from Tables 4.2 and 4.3 that the activation energies for this reaction are much higher than the experimentally accepted value, and hence, we can also conclude that neither pathway is likely to account for cytosine deamination in DNA. Although the barriers for the deamination of cytosine with H_2O are high compared to the experimental value, this mechanism may be important for the catalyzed reaction or for deamination of cytosine in DNA or RNA.

4.3.2 Deamination of Cytosine with OH^-

Since this reaction takes place under physiological conditions (pH 7.4), we have included a detailed study of possible mechanisms for the deamination of cytosine with OH^- . Our results show that in this case, deamination of cytosine could occur via one of two possible pathways, designated as pathways C and D.

The thermodynamic properties for the deamination of cytosine with OH^- are listed in Table 4.4. Two possibilities for the separated products, uracil anion and ammonia, and uracil and amide anion (azanide), were considered as shown in Figure 4.1. Deamination of cytosine with OH^- to form a uracil anion is found to be highly exothermic and exergonic at all levels of theory for the separated species and less so for the formation of the $\text{Ura}^- - \text{NH}_3$ complex. Formation of Ura and NH_2^- is found to be slightly exothermic and exergonic at HF, MP2 and B3LYP levels of theory and to be endothermic and endergonic at B3LYP/6-31+G(d) and G3MP2 levels of theory. This is due to the highly resonance-stabilized uracil anion which has

a highly delocalized negative charge, versus the highly localized negative charge of NH_2^- (an extremely strong base). For the reaction involving separate species, the addition of diffuse functions is shown here to be essential. The B3LYP/6-31+G(d)//B3LYP/6-31G(d) and B3LYP/6-31+G(d)//B3LYP/6-31+G(d) results are in excellent agreement.

4.3.2.1 Deamination of Cytosine with OH^- (Pathway C)

The geometries for the reactant, intermediates, transition states, and product involved in pathway C are shown in Figure 4.7. The relative energies of reactants, intermediates, transition states, and products for pathway C are shown in Figure 4.8.

Pathway C is a five-step mechanism. Deamination of cytosine by OH^- , initially involves deprotonation at the amino group (H_{10}) of cytosine, as shown in Figure 4.7 to form a cytosine anion- H_2O complex. This complex is very stable due to delocalization of the negative charge. Despite extensive attempts, we were not successful in optimizing the geometry of a cytosine- OH^- complex due to the ease of deprotonation of cytosine by OH^- . In the next step, H_2O adds to C_4 , simultaneously adding hydrogen to the imine nitrogen resulting in a tetrahedral intermediate anion (**I1**). This is followed by the formation of three different conformers (**I2**, **I3c** and **I4c**). These structures are very similar, differing mainly in the torsion for H_{15} ($\angle \text{H}_{15}\text{O}_8\text{C}_4\text{N}_3$) and the bond length of $\text{H}_{15}-\text{N}_9$ (Figure 4.7). The reactant, intermediates, transition states, and product illustrated in Figures 4.7 and 4.8 correspond to attack of OH^- on the *si* face (see Figure 4.9) of cytosine. The mirror of this mechanism also exists for attack of OH^- on the *re* face of

cytosine, where the intermediates and transition state structures are enantiomers. In the last step, deamination occurs by intramolecular proton transfer from the hydroxyl group to the amino group to yield a hydrogen-bonded complex of the uracil anion (formal negative charge at N₃) and ammonia. The only other study of cytosine deamination with OH⁻ was reported by Sponer and his co-workers [5]. Their results have shown that the activation energy for this system is 213.4 kJ mol⁻¹ at the B3LYP/6-31G(d) level of theory using the PCM model. The potential energy surface for this system shows that the intermediate and the second transition state have the same energy relative to the reactant (Cyt⁻ - H₂O) complex. The two transition states identified by the authors are not connected in the reaction mechanism as revealed by IRC analysis. One of the transition states belongs to pathway C for attack of OH⁻ on the *re* face of cytosine and the other to pathway C for attack of OH⁻ on the *si* face of cytosine. For the last step of their mechanism they reported an activation energy of 73.6 kJ mol⁻¹ at the B3LYP/6-31G(d) level of theory compared to our value of 97.0 kJ mol⁻¹ at G3MP2 level of theory.

The activation energies, enthalpies of activation, and free energies of activation for the deamination of cytosine with OH⁻ at the HF, MP2, B3LYP levels of theory using the 6-31G(d) basis set and the G3MP2 level of theory for pathway C are listed in Table 4.5. An activation energy of 213.4 kJ mol⁻¹ [5] was reported for the deamination of cytosine with OH⁻ at the B3LYP/6-31G(d) using the PCM solvation model, which is far from the experimental value. In this study, the activation energies for the first rate-determining step are 199.5, 188.9, 177.3, and 148.0 kJ mol⁻¹ at the HF, MP2, B3LYP, and G3MP2 levels of theory, respectively. The

activation energies for the second rate-determining step are 138.0, 126.5, 92.7, and 97.0 kJ mol⁻¹ at the HF, MP2, B3LYP, and G3MP2 levels of theory, respectively. However, the overall activation energy for this pathway is 203 kJ mol⁻¹. The G3MP2 result for the first rate-determining step is fairly close to the experimental value. This difference could be due in part to the environment of cytosine in DNA and RNA, where it is bonded to a sugar moiety and is hydrogen bonded to other nucleic acid bases.

4.3.2.2 Deamination of Cytosine with OH⁻ (Pathway D)

The structures of the reactant, intermediates, transition states, and product involved in pathway D are shown in Figure 4.10. The relative energies of reactants, intermediates, transition states, and products for pathways D are shown in Figure 4.11. The complete reaction pathway was determined with the help of IRC analysis. Pathway D is a five-step mechanism. The first two-steps are the same as those in pathway C. After the intermediate, **I2**, the reaction bifurcates into two different pathways, pathway C via transition state **TS3c** and pathway D via transition state **TS3d**. In step four of pathway D, intermediate **I3d** is converted, by elimination of NH₃, to a uracil anion tautomer-NH₃ complex via **TS4d**, where the C₄-NH₂ bond breaks and the NH₂ group abstracts the hydrogen atom (H₁₃) from N₁. In the final step, a 1-3 proton shift from the hydroxy group to the sp² nitrogen of the uracil tautomer results in formation of a uracil anion with the negative charge localized at N₁.

The activation energies, enthalpies of activation, and free energies of activation for the

deamination of cytosine with OH^- at the HF, MP2 and B3LYP levels of theory using the 6-31G(d) basis set and the G3MP2 level of theory for pathway D are listed in Table 4.6. The activation energy for the rate-determining step are 218.6, 233.3, 170.4, and 158.0 kJ mol^{-1} at the HF, MP2, B3LYP, and G3MP2 levels of theory, respectively. Again, as for the reaction with water, the MP2/GTMP2Large results, for both the thermochemistry (Table 4.4) and barriers (Tables 4.5 and 4.6), are in good agreement with the G3MP2 values. Except for **TS4d**, where G3MP2 and MP2/GTMP2Large results differ by 24 kJ mol^{-1} , the results are all within 12 kJ mol^{-1} .

In summary, the more likely mechanism for the deamination of cytosine is pathway C with an overall activation energy of 203 kJ mol^{-1} , which is still high compared to the experimental value, whereas pathway D is less likely to be the mechanism for this reaction due to the high barriers.

4. 4 Conclusions

The mechanism for the deamination reaction of cytosine with H_2O and OH^- to produce uracil was investigated using *ab initio* calculations. Optimized geometries of reactants, transition states, intermediates, and products were determined at the RHF/6-31G(d), MP2/6-31G(d), and B3LYP/6-31G(d) levels of theory, and at the B3LYP/6-31+G(d) level for the anions. Single point energies were also determined at B3LYP/6-31+G(d), MP2/GTMP2Large and the G3MP2 levels of theory. Thermodynamic properties (ΔE , ΔH , and ΔG), activation energies, enthalpies and free energies of activation were calculated for each reaction pathway that was investigated. Intrinsic Reaction Coordinate analysis was performed to characterize the transition states on the potential energy surface to obtain the complete reaction pathway. The MP2/GTMP2Large result is overall in better agreement with the G3MP2, which may be an acceptable level of theory for larger systems.

Two pathways for deamination with H_2O were found, a five-step mechanism (pathway A) and a two-step mechanism (pathway B). The activation energy for the rate-determining steps, the formation of the tetrahedral intermediate for pathway A, and the formation of the uracil tautomer for pathway B, are 221.3 and 260.3 kJ mol^{-1} at the G3MP2 level of theory, respectively. The deamination reaction by either pathway is therefore unlikely because of the high barriers that are involved but may be a viable mechanism for catalyzed reactions, i.e., with bisulfite. Two pathways for deamination with OH^- were also found, and both of them are five-step mechanisms. Pathways C and D produce an initial tetrahedral intermediate by adding H_2O

to deprotonated cytosine which then undergoes three conformational changes. The final intermediate dissociates to product via a 1-3 proton shift. Deamination with OH^- , through pathway C, resulted in the lowest activation energy for the first rate-determining step, $148.0 \text{ kJ mol}^{-1}$ at G3MP2 level of theory. However, the overall activation energy for this reaction is 203 kJ mol^{-1} at G3MP2 level of theory.

4.5 References

- (1) Neidle, S. *Oxford HandBook of Nucleic Acid Structure*; Oxford University Press Inc.:New York, **1999**.
- (2) Burgur, A. *A Guide to the Chemical Basis of Drug Design*; Wiley: New York, **1983**.
- (3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, **1983**.
- (4) Yao, L.; Li, Y.; Wu, Y.; Liu, A.; Yan, H. *Biochemistry* **2005**, 44, 5940–5947.
- (5) Sponer, J. E.; Miguel, P. J.; Rodriguez-Santiago, L.; Erxleben, A.; Krumm, M.; Sodupe, M.; Sponer, J.; Lippert, B. *Angew. Chem. Int. Ed.* **2004**, 43, 5396–5399.
- (6) Person, W.B.; Szczepaniak, K.; Szczesniak, M.; Kwiatkowski, J. S.; Hernandez, L.; Czerminski, R. *J. Mol. Struc: THEOCHEM* **1989**, 194, 239.
- (7) Shapiro, R.; Servis, R. E.; Welcher, M. *J. Am. Chem. Soc.* **1970**, 92, 422.
- (8) Shapiro, R.; Klein, R. *Biochemistry* **1967**, 6, 3576.
- (9) Brown, D.; Phillips, J. H. *J. Mol. Biol.* **1965**, 11, 663–671.
- (10) Notari, R. E.; Chin, M. L.; Cardoni, A. *J. Pharm. Sci.* **1970**, 59, 28.
- (11) Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, 98, 6338.
- (12) Shapiro, R.; Klein, R. *Biochemistry* **1966**, 5, 2358.
- (13) Peng, W.; Shaw, B. R. *Biochemistry* **1996**, 35, 10172–10181.
- (14) Chen, H.; Shaw, B. R. *Biochemistry* **1994**, 33, 4121–4129.
- (15) Bobek, M.; Cheng, Y. C.; Bloch, A. *J. Medicinal Chemistry* **1978**, 21, 597.

- (16) Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. *Biochemistry* **1990**, 29, 2532–2537.
- (17) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1989**, 111, 2308–310.
- (18) Florián, J.; Baumruk, V.; Leszczyński, J. *J. Phys. Chem.* **1996**, 100, 5578–5589.
- (19) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, 98, 3161–3164.
- (20) Ji, Y. J.; Xia, Y. Y.; Zhao, M. W.; Huang, B. D.; Li, F. *J. Mol. Struc.: THEOCHEM* **2005**, 723, 123–129.
- (21) Wysokinski, R.; Bienko, D. C.; Michalska, D.; Zeegers-Huyskens, T. *Chem. Phys.* **2005**, 315, 17–26.
- (22) Florián, J.; Leszczyński, J. *J. Am. Chem. Soc.* **1996**, 118, 3010–1017.
- (23) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem.* **1996**, 100, 1965–1974.
- (24) Hobza, P.; Šponer, J.; Polášek, M. *J. Am. Chem. Soc.* **1995**, 117, 792–798.
- (25) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, 110, 8319–330.
- (26) Smets, J.; Jalbout, A. F.; Adamowicz, L. *Chem. Phys. Lett.* **2001**, 342, 342–346.
- (27) Hou, X.-J.; Nguyen, M. T. *Chem. Phys.* **2005**, 310, 1–9.
- (28) Fülcher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, 117, 2089–2095.
- (29) Danilov, V. I.; Slyusarchuk, O. N.; Alderfer, J. L. *Photochemistry and Photobiology* **1994**, 59, 125–129.
- (30) Kwiatkowski, J. S.; Leszczyński, J. *J. Phys. Chem* **1996**, 100, 941–953..
- (31) Danilov, V. I.; Les, A.; Alderfer, J. L. *Polish J. Chem.* **2001**, 75, 1039–1049.

- (32) Monajjemi, M.; Ghiasi, R.; Ketabi, S.; Passdar, H.; Mollaamin, F. *J. Chemical Research* **2004**, 11–18.
- (33) Zhanpeisov, N. U.; Leszczyński, J. *J. Phys. Chem. B* **1998**, 102, 9109–9118.
- (34) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Comput. Chem.* **1996**, 17, 841–850.
- (35) Estrin, D. A.; Paglieri, L.; Corongiu, G. *J. Phys. Chem.* **1994**, 98, 5653–5660.
- (36) Scanlan, M. J.; Hillier, I. H. *J. Am. Chem. Soc.* **1984**, 106, 3737–3745.
- (37) Chandra, A. K.; Nguyen, M. T.; Zeegers-Huyskens, T. *J. Mol. Struct.* **2000**, 519, 1–11.
- (38) Chandra, A. K.; Michalska, D.; Wysokinsky, R.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **2004**, 108, 9593–9600.
- (39) Broo, A.; Holmén, A. *J. Phys. Chem. A* **1997**, 101, 3589–600.
- (40) Monajjemi, M.; Ghiasi, R.; Abedi, A. *Theor. Inorg. Chem.* **2005**, 50, 435–441.
- (41) Burda, J.; Šponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem. B* **1997**, 101, 9670–9677.
- (42) Šponer, J.; Burda, J. V.; Sabat, M.; Leszczyński, J.; Hobza, P. *J. Phys. Chem. A* **1998**, 102, 5951–5957.
- (43) Leś, A.; Adamowicz, L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, 93, 4001–4005.
- (44) Alemán, C. *Chem. Phys.* **2000**, 253, 13–19.
- (45) Civcir, P. Ü. *J. Mol. Struct.: THEOCHEM* **2000**, 532, 157–169.
- (46) Gould, I. R.; Green, D. V. S.; Young, P.; Hillier, I. H. *J. Org. Chem.* **1992**, 57, 4434–4437.
- (47) Morpurgo, S.; Bossa, M.; Morpurgo, G. O. *Adv. Quant. Chem.* **2000**, 36, 169.
- (48) Sambrano, J. R.; Souza, A. R.; Queralt, J. J.; Andrés, J. *Chem. Phys. Lett.* **2000**, 317,

437–443.

(49) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, 118, 6811–6821.

(50) Topal, M. D.; Fresco, J. R. *Nature* **1976**, 263, 285.

(51) Fogarasi, G. *J. Phys. Chem. A* **2002**, 106, 1381.

(52) Sobolewski, A. L.; Adamowicz, L. *J. Chem. Phys.* **1995**, 102, 5708.

(53) Fogarasi, G.; Szalay, P. G. *Chem. Phys. Lett.* **2002**, 356, 383.

(54) Prado, M. A. S.; Garcia, E.; Martins, J. B. L. *Chem. Phys. Lett.* **2006**, 418, 264–267.

(55) Leszczyński, J. *J. Phys. Chem.* **1992**, 96, 1649–1653.

(56) Glaser, R.; Rayat, S.; Lewis, M.; Son, M.-S.; Meyer, S. *J. Am. Chem. Soc.* **1999**, 121, 6108–6119.

(57) Lindahl, T.; Nyberg, B. *Biochemistry* **1974**, 13, 3405–3410.

(58) Duncan, B. K.; Miller, J. H. *Nature* **1980**, 287, 560–561.

(59) Flinn, C.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2003**, 107, 11174–11181.

(60) (a) Poirier, R. A. *MUNgauss*, (Fortran 90 version); Chemistry Department, Memorial University of Newfoundland, St. John's, Newfoundland, A1B 3X7 Canada. With contributions from Brooker, M.; Bungay, S. D.; El-Sherbiny, A.; Gosse, T.; Keefe, D.; Pye, C. C.; Reid, D.; Shaw, M.; Wang Y.; Xidos, J. (b) Colonna, F.; Jolly, L.-H.; Poirier, R. A.; Angyan, J.; and Jansen, G. *Comput. Phys. Commun.* **1994**, 81, 293–317.

(61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr. T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.

A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03*, Revision B.05, Gaussian, Inc.: Pittsburgh PA, **2003**.

(62) Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A. *Can. J. Chem.* **2005**, 83, 2082–2090.

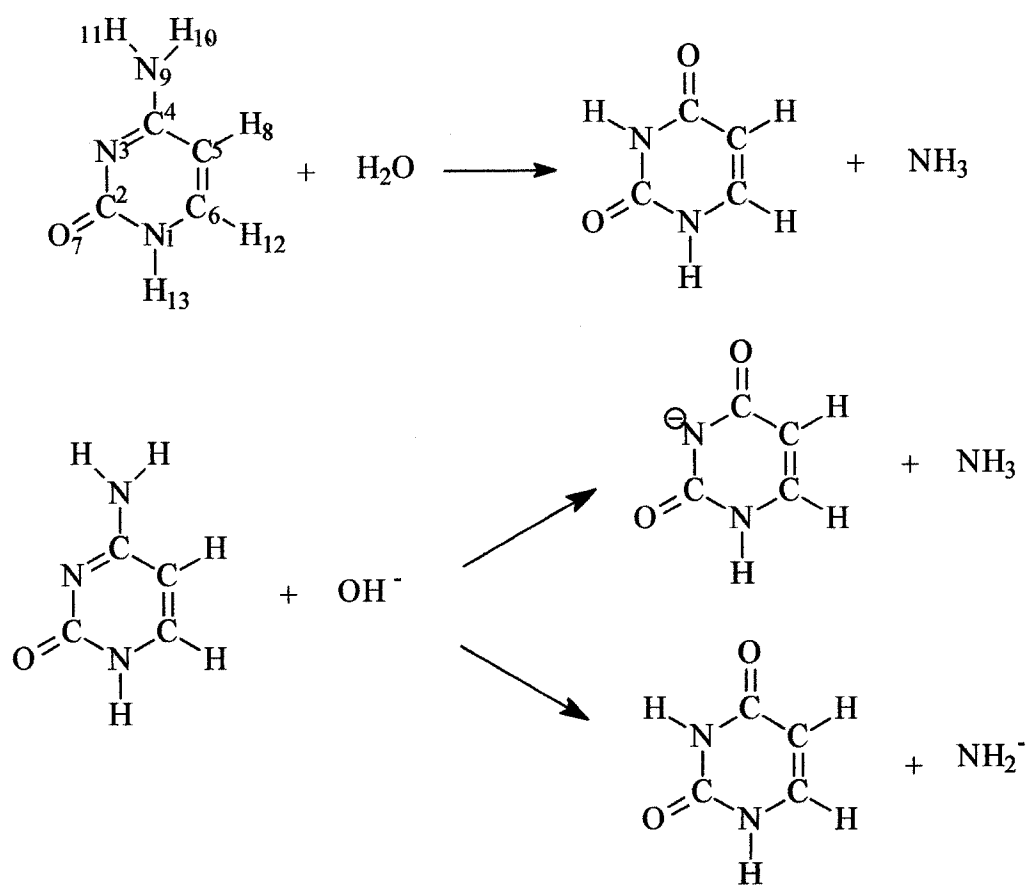


Figure 4.1: Deamination of cytosine with H_2O , deamination of cytosine with OH^- with two possible products, and atom labeling in the amino-oxo tautomer of cytosine.

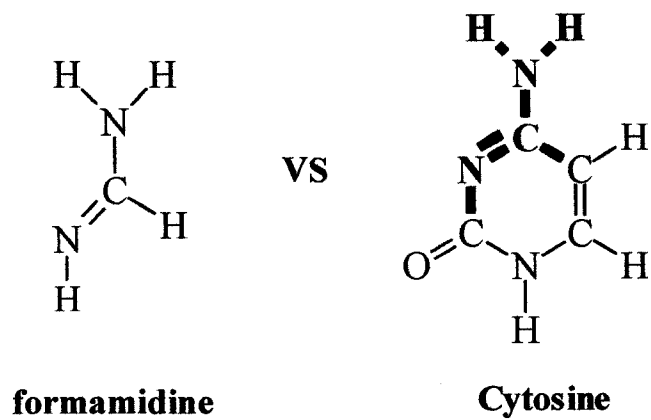


Figure 4.2: Comparison of formamidine (*E*-isomer) and cytosine structure

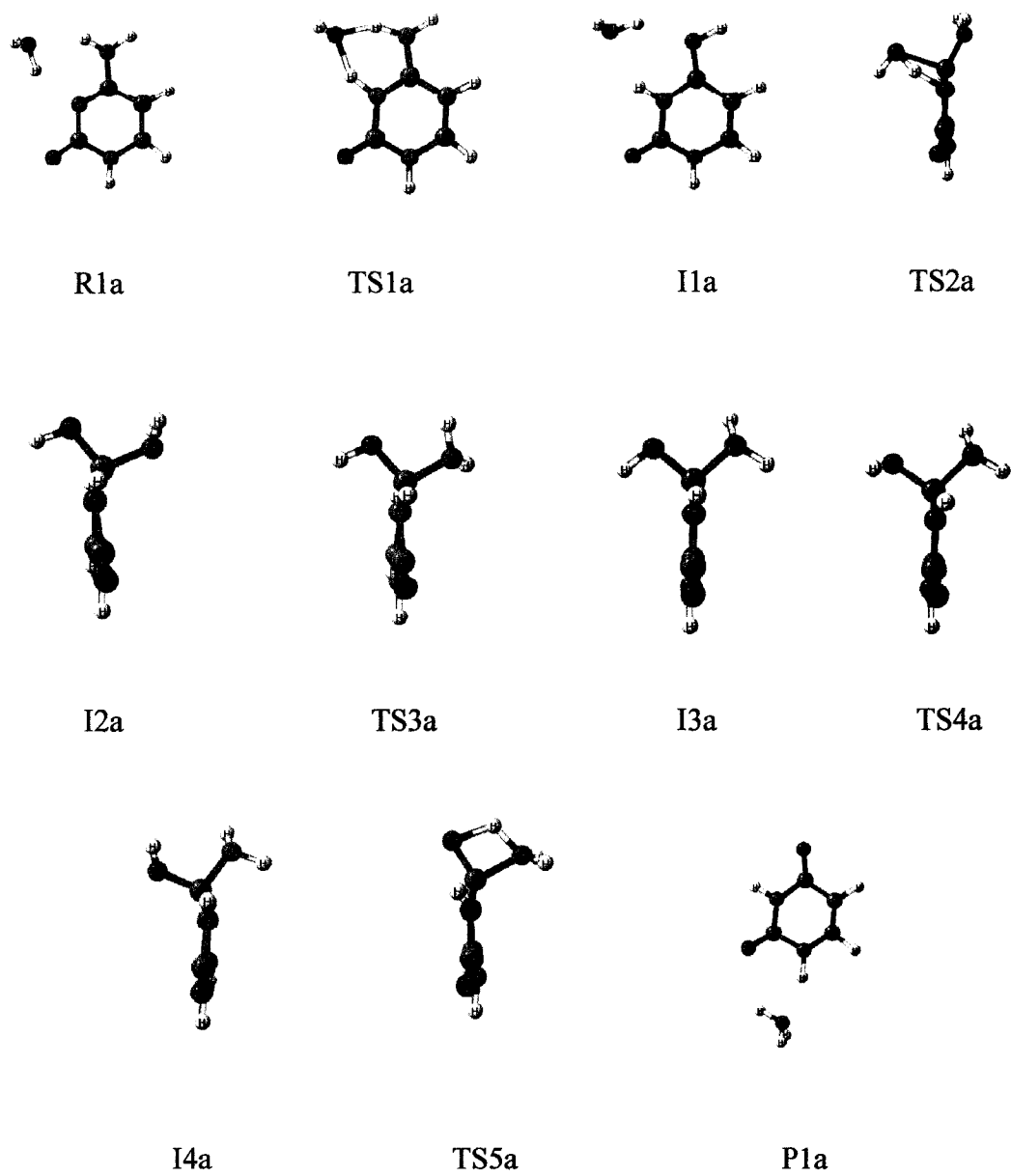


Figure 4.3: Deamination of cytosine with one water molecule to form the Ura-NH₃ complex (pathway A).

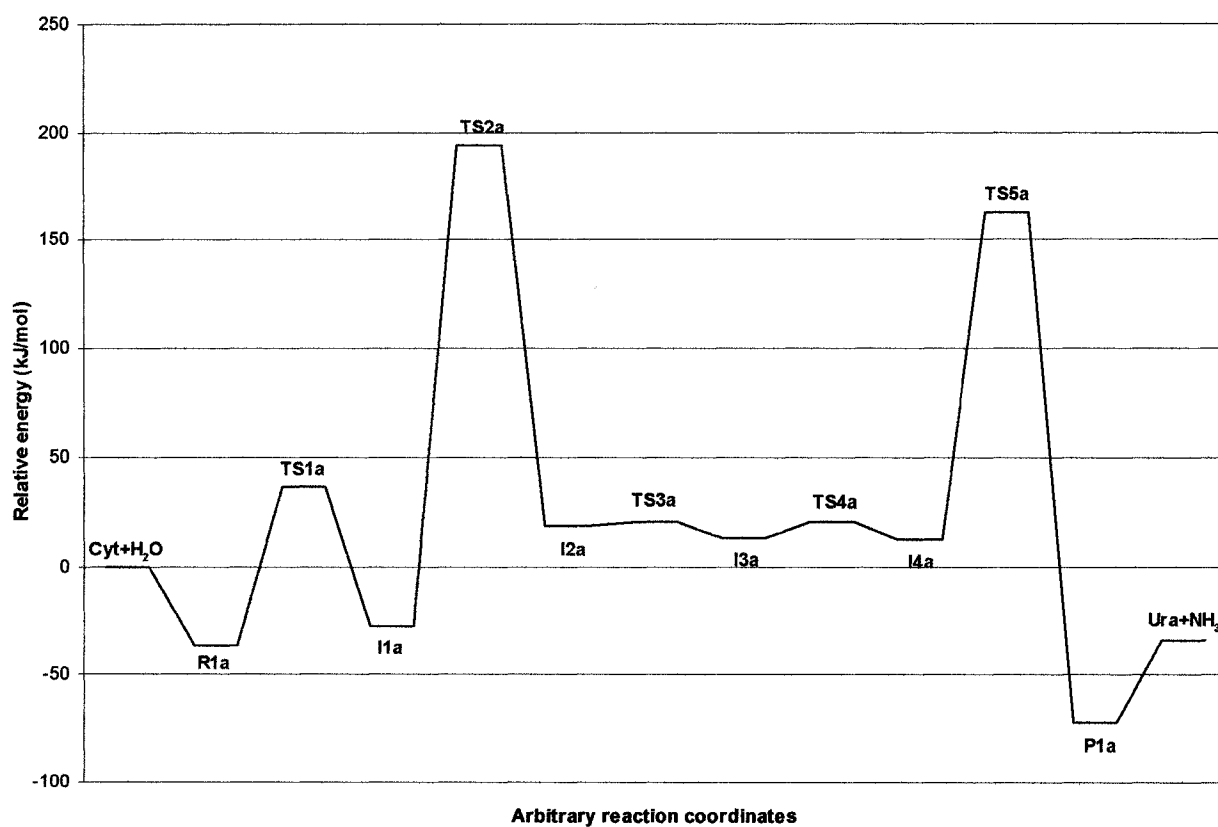


Figure 4.4: Reaction pathway for the deamination of cytosine with one water molecule at the G3MP2 level of theory (pathway A)

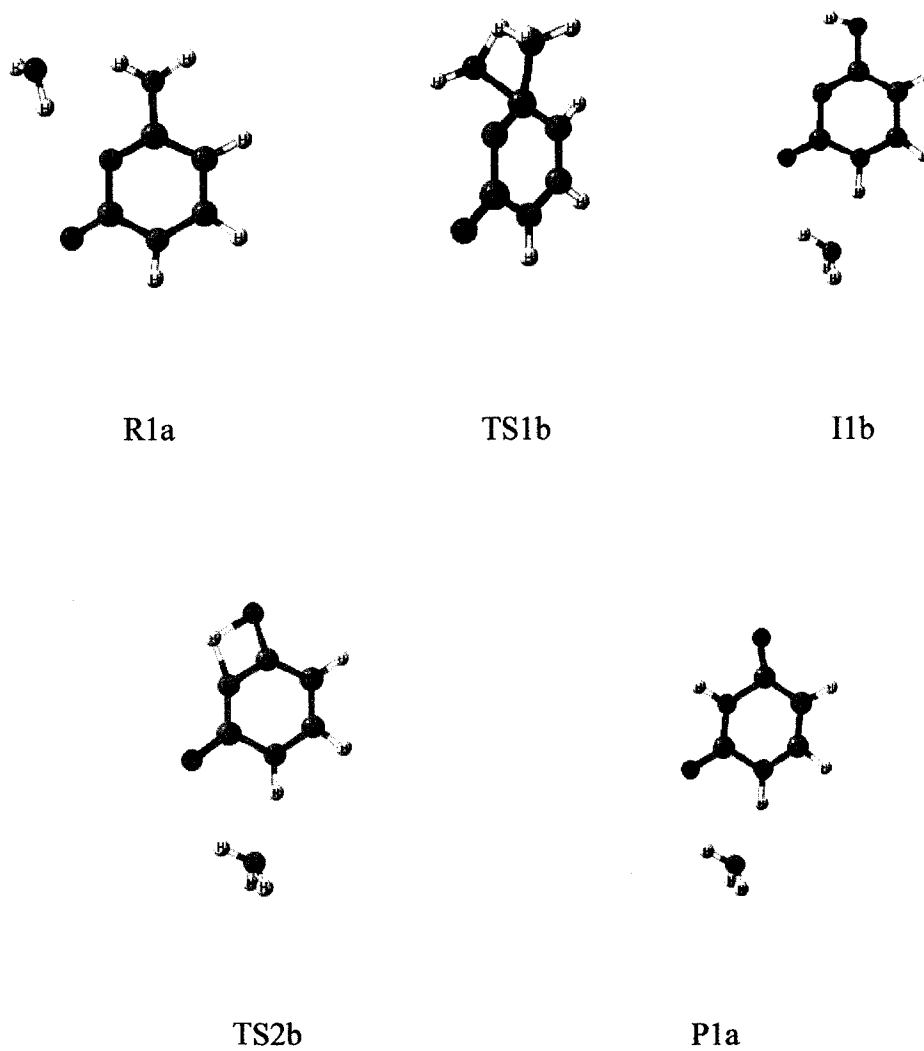


Figure 4.5: Deamination of cytosine with one water molecule to form the Ura-NH₃ complex (pathway B).

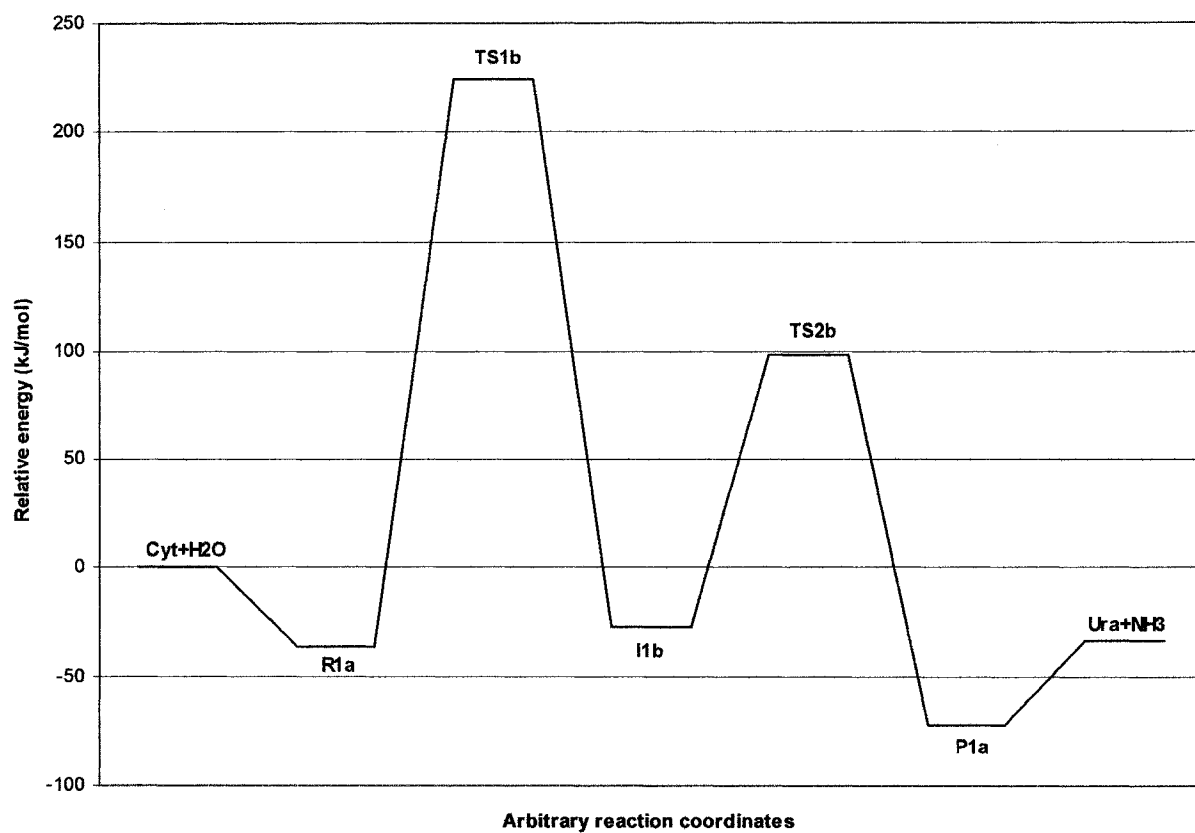


Figure 4.6: Reaction pathway for the deamination of cytosine with one water molecule at the G3MP2 level of theory (pathway B).

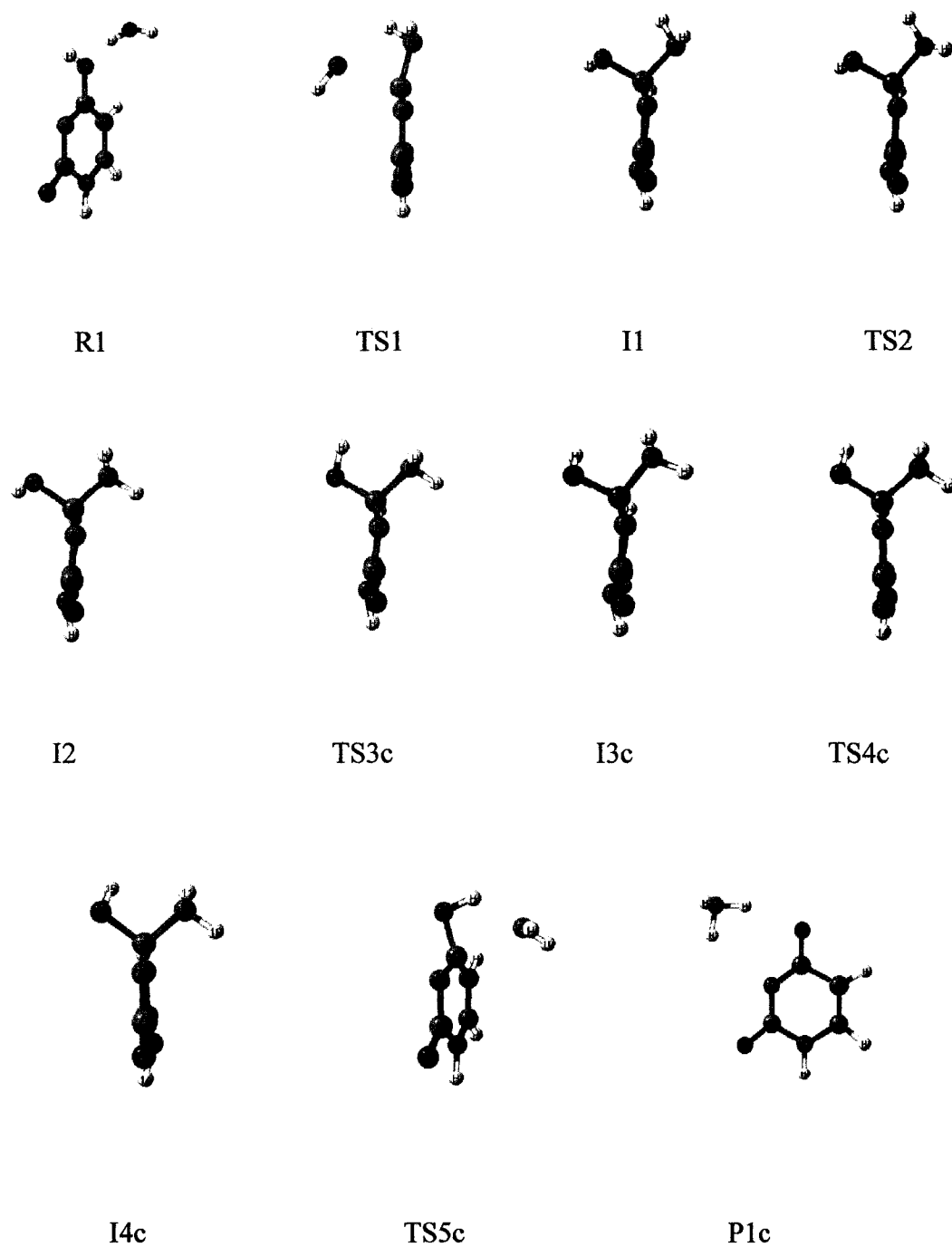


Figure 4.7: Deamination of cytosine with OH^- on the *si* face (see Figure 4.9) to form the Ura anion- NH_3 complex (pathway C).

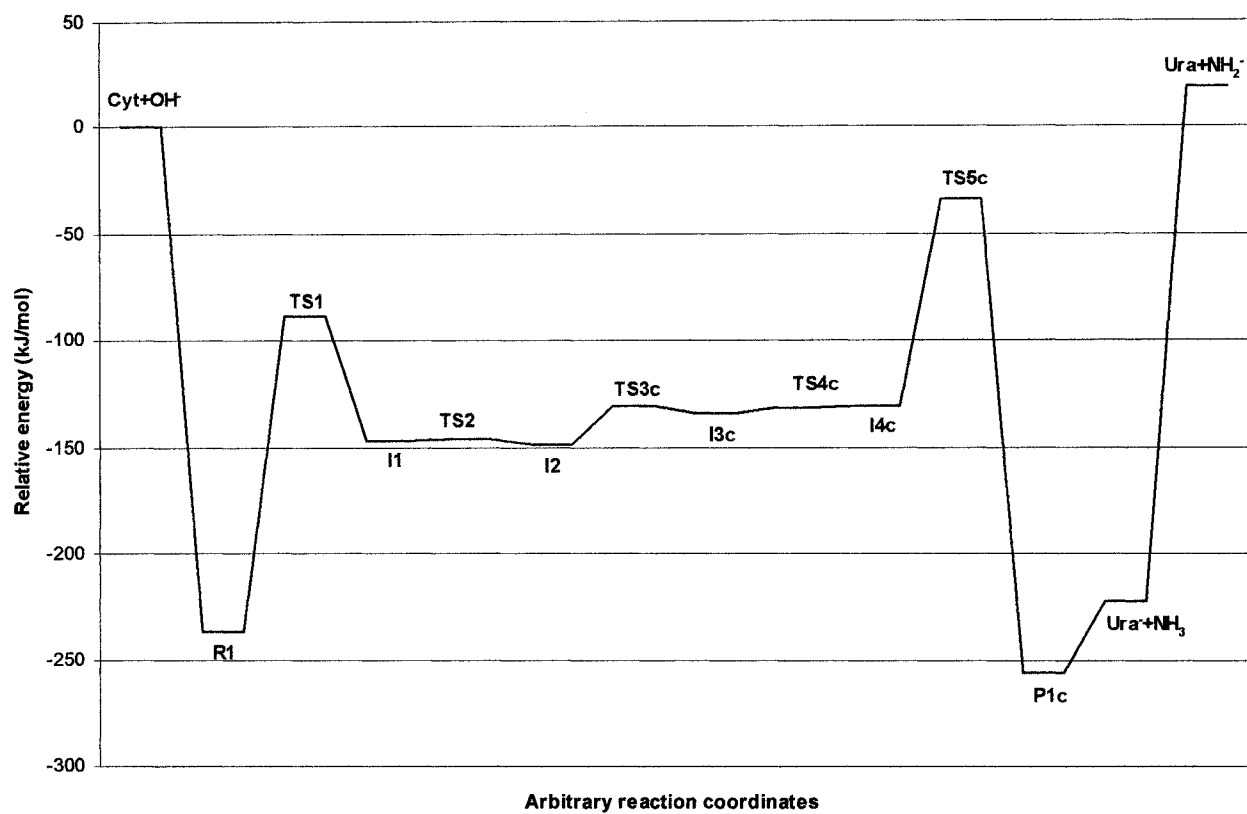


Figure 4.8: Reaction pathway for the deamination of cytosine with OH⁻ at the G3MP2 level of theory (pathway C)

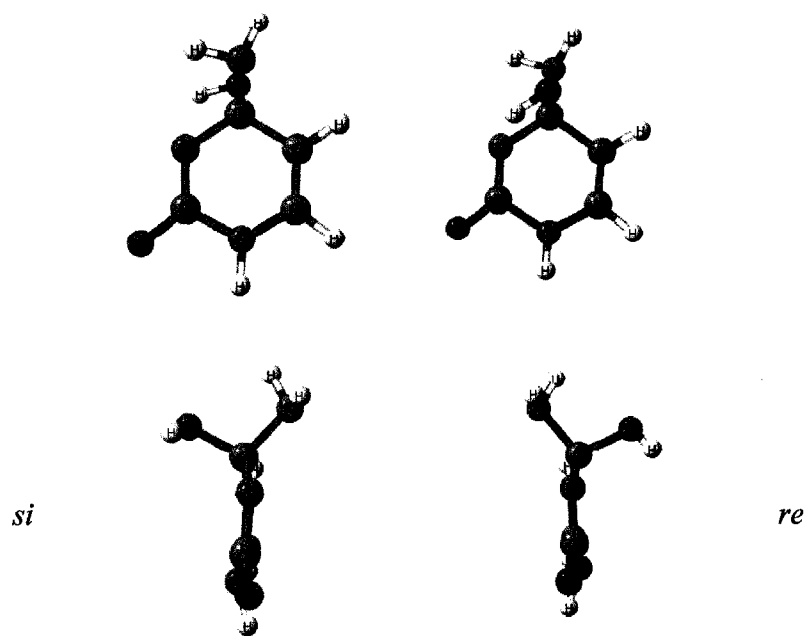


Figure 4.9: Two enantiomers of the intermediate structure (I1) corresponding to attack of OH^- on the *si* face and on the *re* face of the cytosine ring.

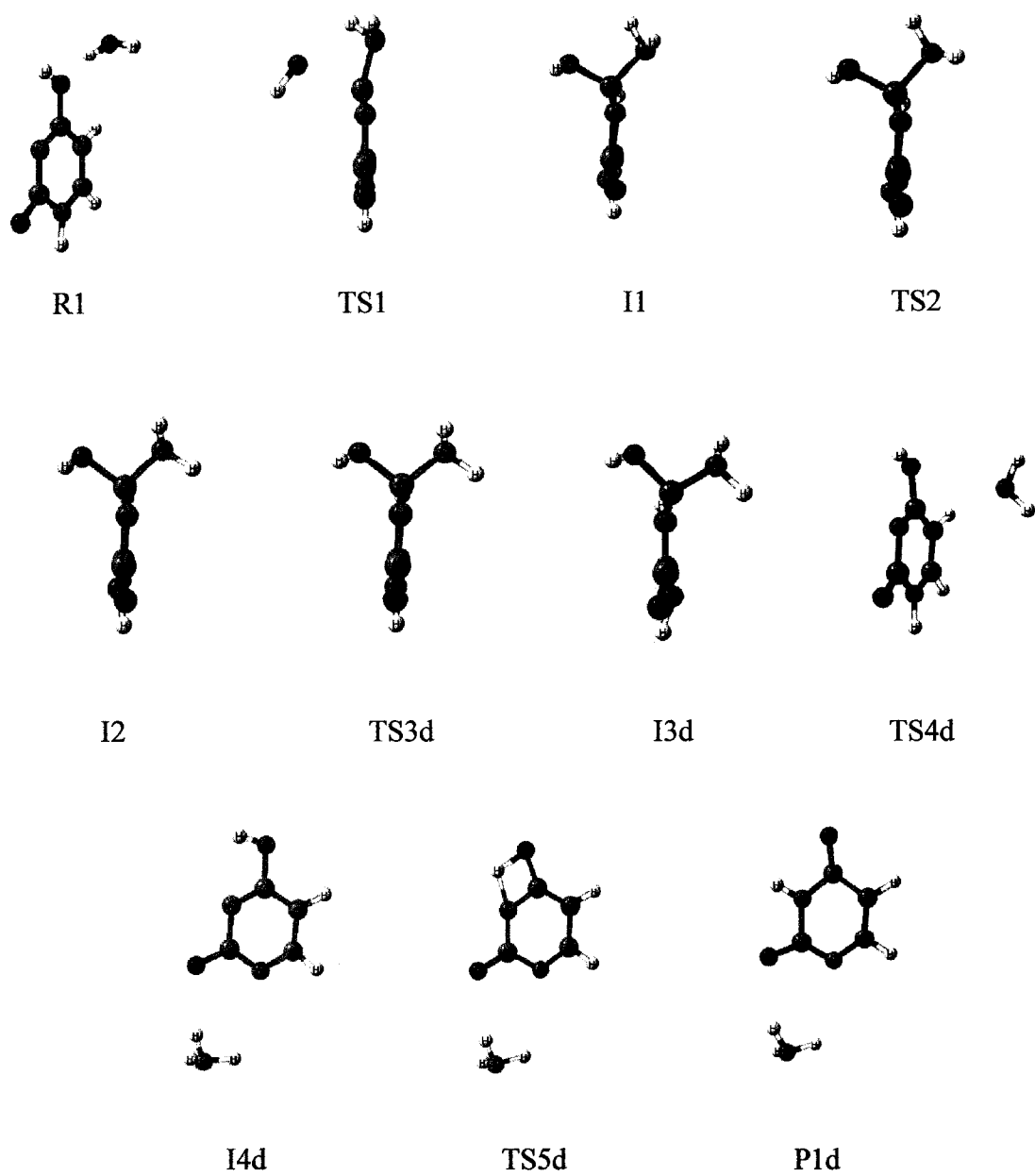


Figure 4.10: Deamination of cytosine with OH^- to form the Ura anion- NH_3 complex (pathway D).

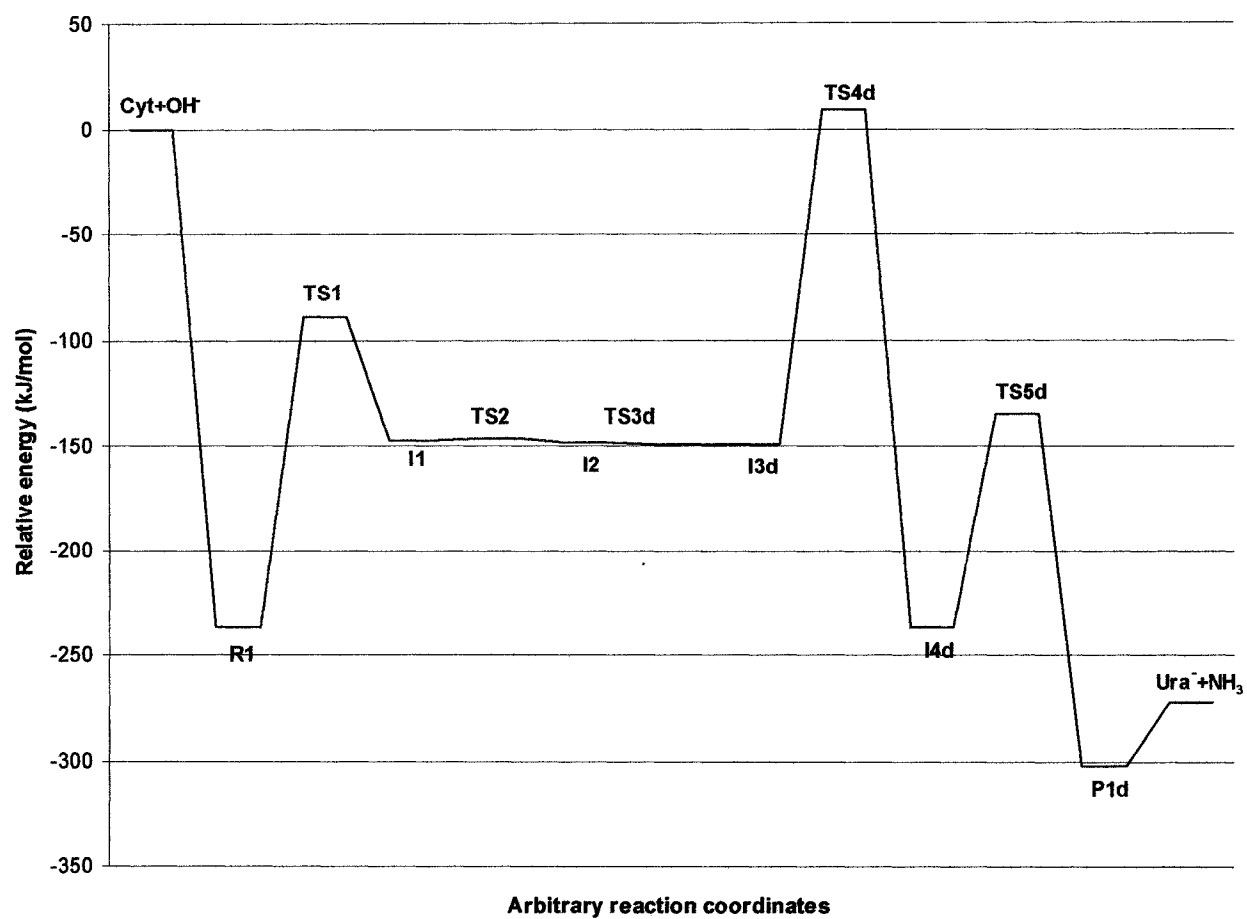


Figure 4.11: Reaction pathway for the deamination of cytosine with OH⁻ at the G3MP2 level of theory (pathway D)

TABLE 4.1: Thermodynamic properties for the deamination of cytosine with one water molecule (in kJ mol⁻¹) at 298.15 K.

	HF/6-31G(d)	MP2/6-31G(d)	B3LYP/6-31G(d)	G3MP2
ΔE^a	-72.3	-62.3	-64.9	-35.3 (-36.2) ^c
ΔH^a	-69.1	-58.8	-62.2	-35.3
ΔG^a	-70.9	-60.5	-63.6	-37.0
ΔE^b	-75.0	-64.8	-70.7	-34.1 (-36.7) ^c
ΔH^b	-71.0	-60.6	-66.7	-35.5
ΔG^b	-70.0	-60.7	-66.1	-34.2

^a For the (Cyt/H₂O complex \Rightarrow Ura/NH₃ complex).

^b For the separated species (Cyt + H₂O \Rightarrow Ura + NH₃).

^c The values in parentheses are for the MP2/GTMP2Large level of theory.

TABLE 4.2: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with one water molecule (in kJ mol⁻¹) at 298.15 K (Pathway A).

	HF/6-31G(d)	MP2/6-31G(d)	B3LYP/6-31G(d)	G3MP2
ΔE_{act} , TS1	138.8	85.3	73.5	72.6 (76.6) ^a
ΔH^\ddagger , TS1	119.3	64.0	53.4	67.8
ΔG^\ddagger , TS1	130.5	72.5	61.6	79.6
ΔE_{act} , TS2	271.3	240.2	223.2	221.3 (223.7) ^a
ΔH^\ddagger , TS2	264.2	226.2	209.8	218.8
ΔG^\ddagger , TS2	270.1	233.1	213.9	225.0
ΔE_{act} , TS3	5.0	5.7	1.8	1.7(2.7) ^a
ΔH^\ddagger , TS3	2.3	3.1	-0.7	0.5
ΔG^\ddagger , TS3	4.1	4.1	1.7	2.5
ΔE_{act} , TS4	8.8	10.6	10.0	6.6 (9.2) ^a
ΔH^\ddagger , TS4	5.6	7.1	6.8	5.4
ΔG^\ddagger , TS4	7.8	8.8	9.5	7.9
ΔE_{act} , TS5	210.6	157.8	145.9	150.1 (156.7) ^a
ΔH^\ddagger , TS5	194.9	141.8	129.3	149.4
ΔG^\ddagger , TS5	196.0	141.1	128.9	150.6

^a The values in parentheses are for the MP2/GTMP2Large level of theory.

TABLE 4.3: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with one water molecule (in kJ mol⁻¹) at 298.15 K (Pathway B).

	HF/6-31G(d)	MP2/6-31G(d)	B3LYP/6-31G(d)	G3MP2
ΔE_{act}, TS1	338.8	258.2	260.1	260.3 (258.3) ^a
ΔH^\ddagger, TS1	329.3	246.5	249.0	256.2
ΔG^\ddagger, TS1	339.6	254.8	256.7	267.1
ΔE_{act}, TS2	186.6	131.8	129.7	125.7 (132.4) ^a
ΔH^\ddagger, TS2	172.0	118.4	116.2	125.2
ΔG^\ddagger, TS2	172.3	118.7	116.4	125.7

^a The values in parentheses are for the MP2/GTMP2Large level of theory.

TABLE 4.4: Thermodynamic properties for the deamination of cytosine with OH⁻ (in kJ mol⁻¹) at 298.15 K.

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d) ^a	G3MP2
Cyt / H₂O complex → Ura⁻ (N₃) / NH₃ complex^b					
ΔE	-53.4	-40.3	-45.5	-33.5 (-35.0)	-18.9 (-15.1) ^c
ΔH	-53.1	-37.9	-43.7	-31.2	-17.1
ΔG	-58.5	-40.8	-48.7	-37.0	-22.6
Cyt + OH⁻ → Ura⁻ (N₃) + NH₃^b					
ΔE	-328.2	-339.6	-347.8	-229.9 (-229.8)	-223.0 (-224.9) ^c
ΔH	-325.2	-337.6	-346.0	-229.3	-223.3
ΔG	-329.4	-344.5	-351.1	-233.8	-227.1
Cyt + OH⁻ → Ura + NH₂⁻					
ΔE	-11.6	-2.9	-21.0	21.2 (23.4)	18.5 (25.7) ^c
ΔH	-15.5	-6.7	-24.9	18.0	18.3
ΔG	-18.5	-10.8	-28.3	15.3	15.7
Cyt / H₂O complex → Ura⁻ (N₁) / NH₃ complex^b					
ΔE	-115.4	-97.3	-103.5	-87.8 (-89.1)	-65.5 (-64.3) ^c
ΔH	-114.1	-93.0	-100.2	-84.1	-64.1
ΔG	-119.1	-94.4	-104.5	-88.3	-69.2
Cyt + OH⁻ → Ura⁻ (N₁) + NH₃^b					
ΔE	-395.7	-402.8	-412.0	-288.5 (-288.3)	-272.6 (-278.0) ^c
ΔH	-391.5	-398.7	-408.4	-286.4	-273.7
ΔG	-393.8	-402.1	-411.2	-288.6	-275.5

^a The values in bracket are for B3LYP/6-31+G(d)// B3LYP/6-31G(d).

^b Ura⁻ (N₁) and Ura⁻ (N₃) for deprotonation at N₁ and N₃ respectively.

^c The values in parentheses are for the MP2/GTMP2Large level of theory.

TABLE 4.5: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway C).

	HF/6-31G(d)	MP2/6-31G(d)	B3LYP/6-31G(d)	G3MP2
ΔE_{act}, TS1	199.5	188.9	177.3	148.0 (153.8) ^a
ΔH^\ddagger, TS1	194.6	184.5	172.5	145.5
ΔG^\ddagger, TS1	204.6	195.8	180.3	155.7
ΔE_{act}, TS2	1.7	2.5	0.9	1.0 (1.8) ^a
ΔH^\ddagger, TS2	-0.5	0.1	-1.4	-0.7
ΔG^\ddagger, TS2	2.4	2.5	2.0	2.6
ΔE_{act}, TS3c	26.1	24.7	21.5	17.5 (19.9) ^a
ΔH^\ddagger, TS3c	23.0	21.6	18.4	16.3
ΔG^\ddagger, TS3c	25.0	23.4	19.9	18.6
ΔE_{act}, TS4c	3.0	6.1	8.2	2.7 (2.0) ^a
ΔH^\ddagger, TS4c	-0.1	2.9	4.9	1.2
ΔG^\ddagger, TS4c	3.2	5.0	3.8	4.8
ΔE_{act}, TS5c	138.0	126.5	92.7	97.0 (108.5) ^a
ΔH^\ddagger, TS5c	125.0	110.2	79.9	97.8
ΔG^\ddagger, TS5c	124.4	106.2	77.2	97.2

^a The values in parentheses are for the MP2/GTMP2Large level of theory.

TABLE 4.6: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with OH⁻ (in kJ mol⁻¹) at 298K (Pathway D).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	G3MP2
ΔE_{act}, TS1	199.5	188.9	177.3	148.0 (153.8) ^a
ΔH^\ddagger, TS1	194.6	184.5	172.5	145.5
ΔG^\ddagger, TS1	204.6	195.8	180.3	155.7
ΔE_{act}, TS2	1.7	2.5	0.9	1.0 (1.8) ^a
ΔH^\ddagger, TS2	-0.5	0.1	-1.4	-0.7
ΔG^\ddagger, TS2	2.4	2.5	2.0	2.6
ΔE_{act}, TS3d	0.4	2.4	0.5	-0.3 (0.4) ^a
ΔH^\ddagger, TS3d	-2.5	-0.8	-2.3	-1.9
ΔG^\ddagger, TS3d	1.1	1.3	1.2	1.9
ΔE_{act}, TS4d	218.6	233.3	170.4	158.0 (182.3) ^a
ΔH^\ddagger, TS4d	204.1	218.9	157.5	161.4
ΔG^\ddagger, TS4d	196.2	211.2	151.6	153.2
ΔE_{act}, TS5d	157.6	107.7	103.6	101.9 (108.9) ^a
ΔH^\ddagger, TS5d	143.5	94.8	90.7	101.4
ΔG^\ddagger, TS5d	143.6	95.2	91.3	101.7

^a The values in parentheses are for the MP2/GTMP2Large level of theory.

CHAPTER 5

Mechanisms for the Deamination Reaction of Cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$: A Computational Study⁶

5.1 Introduction

A detailed knowledge of the structure of cytosine (Cyt) and its nucleoside as well as the tautomerism of nucleic acid bases is an important prerequisite in understanding the molecular basis underlying their biological and medicinal functions. The genetic code can be affected by tautomers of cytosine which can be recognized as other nucleic acid bases (i.e., uracil or thymine); in particular, rare tautomers are thought to induce alterations in the normal base pairing leading to spontaneous mutations and errors in the genetic code, in the DNA or RNA helices [1]. Cytosine occurs naturally in all nucleic acids, including both DNA and RNA. It is chemically bound to the sugar moiety and interacts with other nucleic acid bases via hydrogen

⁶ Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A., Submitted to the *J. Chem. Inf. Model.* on August 29th, 2007.

bonds, most frequently with guanine [2]. For a more detailed review, see our previous work [3] and the references cited therein.

Mutations or changes to the nucleotide sequences of DNA can arise in a number of ways. Mistakes can be made during DNA replication that result in the incorporation of an incorrect base. The chemical nature of any given base can be altered either by environmental or chemical means. Once altered, these changes may then be propagated by further DNA replication. Finally, large scale changes can sometimes occur in the form of DNA insertions and/or deletions. Spontaneous mutations can arise as a result of chemical changes to individual bases in DNA. One such chemical change is the conversion of cytosine to uracil (Ura) which is classified as a deamination reaction (the loss of an amino group from a tetrahedral carbon). The hydrolytic deamination reaction of cytosine to yield uracil is shown in Figure 5.1. Uracil is found in RNA and can base pair with adenine. It has firmly been established that deamination of the DNA base cytosine is an extremely rare event under normal physiological conditions (40-100 deaminations in the human genome per day, pH 7.4) although the rate of deamination can be significantly increased in the presence of various reagents such as NO, HNO₂ and bisulfite.

If uracil is found in DNA, it poses a very serious problem. The cell, however, has a specific enzyme to remove it from DNA, called DNA uracil-N-glycosylase. The uracil formed by cytosine deamination is potentially mutagenic, changing the coding information during DNA replication and RNA transcription, resulting in altered base pairs in the genome [4].

The deamination of cytosine, in particular, and a number of its derivatives have been the subject of many experimental studies [5-15]. Federico et al. [15] were able to determine the rate constant of cytosine deamination for single- and double-stranded DNA at physiologically relevant conditions (37 C° and pH 7.4) by a sensitive genetic assay. Their measured rate constants for single and double- stranded DNA are 1×10^{-10} and $7 \times 10^{-13} \text{ s}^{-1}$, respectively, with an activation energy of $117 \pm 4 \text{ kJ mol}^{-1}$. This value agrees well with the previous value of 121 kJ mol^{-1} by Lindahl and Nyberg [16] obtained over a 25 C° temperature range. As a part of their investigation of the deamination of 1-methylcytosine and 1,5-dimethylcytosine with Pt^{II} complex, both experimentally and using DFT calculations, Sponer et al. [17] also studied the deamination of cytosine with OH^- using the PCM model to account for solvation effects. However, their reported activation energy barrier ($213.4 \text{ kJ mol}^{-1}$) is not very close to the experimentally accepted value ($117 \pm 4 \text{ kJ mol}^{-1}$) [15, 16] and the mechanism reported in their paper differs in a number of ways from the one reported in our previous work [3]. Rayat et al. [18], studied the nitrosative deamination of cytosine in DNA experimentally. They provided a mechanistic hypothesis for nitrosative cytosine deamination which involves a pyrimidine ring-opened intermediate and proposed a number of possible reaction channels. More recently, Matsubara et al. [19], studied the catalysis for cytidine deaminase. As a part of their investigation, they also studied the uncatalysed hydrolytic deamination of cytosine with H_2O . They found that the catalytic action of cytidine deaminase is effectively enhanced by the participation of the extra water molecule. Their reported activation energy barrier with one water molecule is $237.4 \text{ kJ mol}^{-1}$ at B3LYP/6-31G(d,p)). We have shown in previous work on

the deamination of cytosine with H_2O , that this reaction is not the most likely mechanism to account for the deamination reaction of cytosine and hence not a model for similar systems, such as cytosine derivatives (Cytidine) [3].

From our previous results, the deamination reaction of formamidine and cytosine [3, 20] gave a useful starting point for the study of the mechanism of the deamination reaction of cytosine, providing good initial guesses for the structure of transition states for the proposed mechanism in this paper. Our previous study [3] has shown that the deamination of cytosine with H_2O has a high activation energy ($221.3 \text{ kJ mol}^{-1}$ for pathway A and $260.3 \text{ kJ mol}^{-1}$ for pathway B). The activation energy for the deamination reaction with OH^- was very high, with an overall activation energy of $203.1 \text{ kJ mol}^{-1}$ at the G3MP2 level of theory, compared to the experimental value ($117 \pm 4 \text{ kJ mol}^{-1}$). Since this reaction takes place at physiological condition, pH 7.4, which is slightly basic, and the activation energy of the first rate-determining step for the deamination with OH^- was fairly close to the experimental value. Federico et al. [15] suggested that the probability of deamination reaction would increase by any process that would facilitate OH^- attack on the C_4 residue of cytosine. Since most proton transfers are mediated by water, one must consider the role of water molecules in the proton transfer. OH^- forms a strong hydrogen bond with H_2O and a stable complex. Computational studies have shown that interaction with water changes the relative energies of the cytosine tautomers, the canonical tautomer being better hydrated than the other tautomers [21]. The mechanism for the deamination of cytosine in DNA is unknown and the experimental activation energy was

determined for cytosine deamination in single-stranded DNA. For these reasons, we have performed a detailed study of possible mechanisms for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$ in which the water molecules act as a solvent (and do not directly participate in the proton transfer process), as well as when they mediate the hydrogen transfer step.

5.2 Computational Method

All the computations were performed with the Gaussian03 suite of programs [22]. The geometries of all reactants, transition states, intermediates, and products were fully optimized at the restricted HF, MP2, and B3LYP levels of theory using the 6-31G(d) basis set and at B3LYP/6-31+G(d). Single point energies were determined at G3MP2 and the MP2/GTMP2Large levels of theory. We chose the G3MP2 level of theory which is known to give reliable energetics as confirmed from our previous results [23]. The complete reaction pathway for each mechanism discussed in this paper has been verified using Intrinsic Reaction Coordinate (IRC) analysis of all transition states. The structures at the last IRC points in both directions were further optimized, in order to positively identify the reactant and product to which each transition state is connected. A frequency analysis was performed for each stationary point in order to ensure that all minima had no imaginary frequencies on the potential-energy surface and that transition states had a single imaginary frequency.

5.3 Results and Discussion

Cytosine has significant hydrogen bonding abilities; in particular, it possesses both hydrogen bond donor and acceptor groups. For this reason, we have examined a number of possible hydrogen-bonded complexes between H_2O , OH^- , and cytosine. We found that adding H_2O and OH^- to the $\text{N}_7\text{-C}_4\text{-C}_5\text{-C}_6$ side produced higher activation energies than the $\text{N}_7\text{-C}_4\text{-N}_3\text{-C}_2$ side of cytosine (see Figure 5.1). In addition, since it is the $\text{N}_7\text{-C}_4\text{-N}_3\text{-C}_2$ side of cytosine that hydrogen bonds with other nucleic acid bases (normally guanine), it is more sterically hindered [2]. Adding one water molecule stabilized the transition states and destabilized the reactant complex $\text{Cyt}^- \cdots \text{H}_2\text{O}$. Deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$ can follow several possible pathways designated as pathways A \rightarrow D, and F \rightarrow H.

5.3.1 Deamination of Cytosine with $\text{H}_2\text{O}/\text{OH}^-$: Pathways A and B

In our previous work, we found two energetically equivalent pathways for the deamination of cytosine with OH^- that involve OH^- attacks on both faces of cytosine [3]. However, due to lack of symmetry, the extra water molecule now results in two different pathways (A and B) for attack on the two faces of cytosine.

Pathway A:

The geometries for the reactants, intermediates, transition states, and products as well

as their relative energies involved in pathway A (OH^- attack on the left face of cytosine with the carbonyl group facing the observer) are shown in Figure 5.2.

For comparison, Figure 5.2 shows the relative stability of some possible reactant and product structures (not connected mechanistically). This figure shows that $(\text{Cyt}^- + \text{H}_2\text{O} + \text{H}_2\text{O})$ is more stable than $(\text{Cyt} + \text{H}_2\text{O} + \text{OH}^-)$ due to the delocalized negative charge on the cytosine anion compared to OH^- . $(\text{Cyt}^- \cdots \text{H}_2\text{O} + \text{H}_2\text{O})$ is less stable than a complex of cytosine anion with two water molecules $(\text{Cyt}^- \cdots 2\text{H}_2\text{O})$. The relative stabilities for the products also follow similar trends to the reactants.

Deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ ($\text{Cyt} + \text{OH}^- + \text{H}_2\text{O} \rightarrow \text{Ura} + \text{NH}_3 + \text{OH}^-$) closely follows the mechanism for the deamination of cytosine with H_2O and as well as the deamination with OH^- , particularly in relation to the two rate-determining steps. The first rate-determining step involves formation of a tetrahedral intermediate (II), followed by a conformational change, and a second rate-determining step which is a 1-3 proton shift from the hydroxy group to the exocyclic nitrogen atom.

Pathway A has two rate-determining steps. Deprotonation of cytosine occurs easily without forming a $\text{Cyt} \cdots \text{H}_2\text{O} \cdots \text{OH}^-$ complex. The H_{10} immediately transfers to OH^- forming a water molecule and a more stable $(\text{Cyt}^- \cdots 2\text{H}_2\text{O})$ complex which is the reactant for pathways A, B, C, and D. In the first rate-determining step, the addition of a water molecule stabilizes the

hydroxide ion being formed in the transition state (TS1), dropping the activation energy. Nucleophilic attack by the water molecule on the C₄ carbon atom occurs with simultaneous proton transfer from H₂O to the exocyclic imine nitrogen of the cytosine anion to form a tetrahedral intermediate (I1). This is followed by a conformational change to give a second intermediate I2_A in which the water molecule has migrated to be in the same plane as the cytosine ring, with an activation energy of only 2.5 kJ mol⁻¹ at G3MP2 level of theory. Conformers I1 and I2_A only differ with respect to the position of the water molecule, the torsion of H₁₅ (\angle H₁₅O₁₄C₄N₃), and the torsion of the hydrogen atoms (H₉ and H₁₀) which are connected to N₇. See the atom labelling for I1 in Figure 5.2. The conformational change is followed by an intramolecular 1-3 proton shift from the hydroxyl group to the amino group to yield a Ura⁻...H₂O...NH₃ complex. This mechanism is similar to the deamination of cytosine with OH⁻ [3]. However, addition of a single water molecule stabilizes the transition states of both rate-determining steps lowering the overall activation energy by 8.9 kJ mol⁻¹.

Pathway B:

The geometries for the reactant, intermediates, transition states, and product involved in pathway B, along with the relative energies, are shown in Figure 5.3. Pathway B also has two rate-determining steps. As in pathway A, one of the H₂O molecule attacks the C₄ carbon atom with simultaneous proton transfer to the imino group (sp² exocyclic nitrogen) of cytosine producing a tetrahedral intermediate, I1_B. This is followed by several low barrier conformational changes connecting intermediates I1_B and I4_B. In our previous work on the deamination of

cytosine with H_2O and OH^- , we obtained two intermediate conformers with H_2O and three with OH^- . These conformers are different due to the existence of the two functional groups ($-\text{NH}_2$ and $-\text{OH}$) and H_2O in this system resulting in extra degrees of freedom. These structures are very similar, differing mainly in the torsion and the angles of these functional groups. The activation energies for the conformational changes are very small as expected and do not have a significant effect on the mechanism of this reaction. The final step is a 1-3 proton shift from the hydroxy group to the exocyclic nitrogen of the tautomer (**I4_B**) which results in the uracil anion deamination product, see Figure 5.3.

Figure 5.4 shows a comparison of the reaction pathway for the deamination of cytosine with OH^- and with $\text{H}_2\text{O}/\text{OH}^-$ for pathway B. It can be seen from Figure 5.4 that the barriers for deamination with $\text{H}_2\text{O}/\text{OH}^-$ are noticeably lower than the barriers for the deamination with just OH^- , indicating that the water molecule plays an important role in the deamination reaction.

The activation energies, enthalpies of activation, and free energies of activation for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ at HF, MP2, and B3LYP level of theories using 6-31G(d) basis set, B3LYP/6-31+G(d), and G3MP2 levels of theory for both pathways A and B are listed in Table 5.1 and 5.2, respectively. An activation energy of 117 ± 4 and 121 kJ mol^{-1} were reported experimentally for this reaction [15,16]. In our preliminary computational study [3], the activation energies for the two rate-determining steps are 148.0 and 97.0 kJ mol^{-1} resulting in an overall activation energy of $203.1 \text{ kJ mol}^{-1}$, see Figure 5.4. In this study, the

activation energies for the two rate-determining steps are 115.3 and 114.0 kJ mol⁻¹ for pathway A and 115.7 and 115.2 kJ mol⁻¹ for pathway B at G3MP2 level of theory. Therefore, the overall activation energy for pathways A and B are 194.2 and 197.1 kJ mol⁻¹ at G3MP2 and 192.7 and 193.7 kJ mol⁻¹ at B3LYP/6-31+G(d) level of theory, respectively. The MP2/GTMP2Large results for the barriers, are in excellent agreement with the G3MP2 values, differing by no more than 13 kJ mol⁻¹ as shown in Table 5.1 and 5.2. Furthermore, Table 5.1 and 5.2 show that the computationally less expensive B3LYP/6-31+G(d) results are in good agreement with the G3MP2 differing by no more than 14.5 kJ mol⁻¹ in this study.

5. 3. 2 Deamination of Cytosine with H₂O/OH⁻ : Pathways C and D

The overall activation energies for the deamination of cytosine with H₂O/OH⁻ for both pathways A and B are still high compared to the experimental value. In addition, the extra water molecule for pathways A and B did not participate in the 1-3 proton shift. It was previously reported that the formation of uracil (second step) from the tetrahedral intermediate is the rate-determining step [6,24]. For this step, we examined the role of the water molecule as a mediator on the 1-3 proton shift from the -OH group to the -NH₂ group which is one of the rate-determining steps of the catalytic process. It is well known that water can act not only as a solvent but as a catalyst which can donate or accept a proton to promote long range proton transfer [20, 23]. From our previous study on the decomposition reaction of formamidine [23], we found that adding one water molecule catalysed the 1-3 proton shift by forming a cyclic hydrogen-bonded transition state reducing the barrier by 88.4 kJ mol⁻¹ at G3MP2. Addition

of a second water molecule further reduced the barrier by 17.2 kJ mol⁻¹ at G3MP2.

Pathway C:

The geometries for the reactant, intermediates, transition states and product as well as their relative energies involved in pathway C are shown in Figure 5.5. The activation energies, enthalpies of activation, and free energies of activation for the deamination of cytosine with H₂O/OH⁻ at HF, MP2, and B3LYP level of theories using 6-31G(d) basis set, B3LYP/6-31+G(d), and G3MP2 levels of theory for pathway C are listed in Table 5.3. The first transition state for pathway C is the same as pathway A.

Intermediates **I1** and **I2_c** are connected with several conformational changes, which do not have a significant effect on the mechanism of this reaction and hence are not included in the figures. The barrier for the water-mediated 1-3 proton shift is reduced by 45.7 kJ mol⁻¹ at G3MP2 and 53.5 kJ mol⁻¹ at B3LYP/6-31+G(d) levels of theory. The activation energy for this step is 68.3 kJ mol⁻¹ at G3MP2 and 50.2 kJ mol⁻¹ at B3LYP/6-31+G(d). However, the overall activation energy is still somewhat high, 171.1 kJ mol⁻¹, (160.6 kJ mol⁻¹ at B3LYP/6-31+G(d)), see Table 5.4.

Pathway D:

The geometries for the reactant, intermediates, transition states and product as well as their relative energies involved in pathway D are shown in Figure 5.6. The activation energies,

enthalpies of activation, and free energies of activation for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ at HF, MP2, and B3LYP level of theories using 6-31G(d) basis set, B3LYP/6-31+G(d), and G3MP2 levels of theory for pathway D are all listed in Table 5.4.

Of all the pathways investigated, pathway D (Figure 5.6) is the only pathway that can lead to neutral uracil for this system. The first step for pathway D is the same as pathway A. In this case, intermediate **I1** is converted to **I2_D** by conformational changes. Intermediate **I2_D** undergoes a water-mediated 1–3 proton shift, from the -OH to the N₃ atom, to form intermediate **I3_D** through **TS2_D**. The activation energy for **TS2_D** is 19.2 kJ mol⁻¹ at G3MP2, see Table 5.4. At the MP2/6-31G(d) level of theory, intermediate **I3_D** is converted to intermediate **I4_D** through transition state **TS3_D**, where the -NH₂ group is migrating from C₄ to C₂ (the carbonyl carbon) with the negative charge localized at O₈. In this case the IRC analysis resulted in two other possible intermediates, **I4_D(1)** at B3LYP/6-31G(d) and **I4_D(2)** at HF/6-31G(d) and B3LYP/6-31+G(d) levels of theory. From intermediate **I4_D**, a proton transfers from the oxygen of H₂O to the -NH₂ group with simultaneous transfer of a proton from the N₃ atom of uracil, through **TS4_D**, to form a $\text{Ura}^- \cdots \text{H}_2\text{O} \cdots \text{NH}_3$ complex (**P_D**) with the negative charge at N₃. For pathway D (**R** to **P_D**) the overall activation energy is 155.2 and 135.0 kJ mol⁻¹ at G3MP2 and B3LYP/6-31+G(d) levels of theory, respectively.

However, in reality (in DNA or in solution) intermediate **I3_D**, as shown in Figure 5.6, would most likely be protonated at the -NH₂ group to form ammonia resulting in a complex

of Ura...2H₂O...NH₃. In this case the first step is the rate-determining step with an activation energy of 115.3 and 118.4 kJ mol⁻¹ at G3MP2 and B3LYP/6-31+G(d) levels of theory, respectively, which accounts for the observed experimental value.

5.3.3 Deamination of a Tautomer of Cytosine with 2H₂O:

Pathway F

We also considered the possibility of the Cyt⁻...H₂O...H₂O complex being protonated. Computational studies also have predicted that cytosine protonation in aqueous solution or in the gas phase should occur at the N₃ position [25-29] (see Figure 5.1). Protonation occurred at the N₃ site, known to be the more favourable site for protonation of cytosine [29], resulting in a complex of the amino-oxo tautomer of cytosine with two water molecules. This resulted in a new pathway designated as pathway F. The geometries for the reactant, intermediates, transition state and product as well as their relative energies for pathway F are shown in Figure 5.7. Activation energies, enthalpies of activation, and free energies of activation for pathway F at HF, MP2, and B3LYP level of theories using 6-31G(d) basis set, B3LYP/6-31+G(d), and G3MP2 levels of theory are listed in Table 5.5.

Pathway F is a mechanism with a single rate-determining step and several conformational changes connecting intermediates **I1_F** and **I2_F**. The first and last steps are similar to the previous pathways. Interestingly, the barriers for **TS1_F** and **TS2_F** increase for

pathway F compared to **TS1** and **TS2_C** for pathway C from 115.3 to 140.9 kJ mol⁻¹ and from 68.3 to 92.6 kJ mol⁻¹ at G3MP2 level of theory, respectively. The first step is the rate-determining step with a barrier of 140.9 and 129.6 kJ mol⁻¹ at G3MP2 and B3LYP/6-31+G(d) levels of theory, respectively, which is high compared to the experimental value.

5.3.4 Deamination of Cytosine with 2H₂O/OH⁻: Pathways G and H

Since the presence of a single water molecule had a significant effect on the various reaction pathways, we have investigated the effect of a second water molecule on the potential energy surface for pathways C and D. The geometries of the reactants, intermediates, transition states and products as well as their relative energies for pathways G and H are shown in Figure 5.8 and 5.9, respectively. The activation energies, enthalpies of activation, and free energies of activation pathways G and H at HF, MP2, and B3LYP level of theories using 6-31G(d) basis set, B3LYP/6-31+G(d), and G3MP2 levels of theory are listed in Table 5.6 and 5.7, respectively.

The key steps for pathways G and H are similar to the previous pathways C and D. Addition of the second water molecule further reduces the barriers for both rate-determining steps. For pathway G, the product (**P_G**) is a complex of the uracil anion, two water molecules, and ammonia (Ura⁻...2H₂O...NH₃) at MP2/6-31G(d), B3LYP/6-31G(d) and B3LYP/6-31+G(d)

levels of theory with a negative charge at the N₁ atom, which is not possible in DNA and RNA (cytosine is chemically bound to the sugar moiety at N₁ in DNA). However, this structure does not converge at HF/6-31G(d) and instead the (Ura...H₂O...OH⁻...NH₃) complex is formed where the OH⁻ ion is stabilized by hydrogen bonding with the H₂O and NH₃ molecules.

For pathways G and H, we note that adding the second water molecule reduces the activation energy of the first-step (which is the same for both pathways) by 30.5 kJ mol⁻¹ at G3MP2 and by a lesser amount for the other steps, see Table 5.7. This is due to the fact that the second water molecule can both stabilize the transition state and act as a catalyst for this system. The activation energies for the two rate-determining steps are 84.8 and 90.3 kJ mol⁻¹ for pathway G and 84.8 and 55.3 kJ mol⁻¹ for pathway H at G3MP2 level of theory. However, the overall activation energies for pathways G and H are 140.5 and 152.9 kJ mol⁻¹ at G3MP2 (134.6 and 155.6 kJ mol⁻¹ at B3LYP/6-31+G(d)), respectively, see Tables 5.6 and 5.7, which are still high compared to the experimental value. For pathway G intermediate I3_G, similarly to intermediate I3_D, would most likely be protonated at the -NH₂ group. The first step would then be the rate-determining step with an activation energy of only 93 kJ mol⁻¹ at G3MP2. This lower barrier may account for the difference between the deamination of free cytosine versus cytosine in single-stranded DNA.

5.3.5 Thermodynamic Properties and Relative Stabilities

Table 5.8 shows the thermodynamic properties and relative stabilities of several species investigated. The deamination reaction of cytosine to form uracil and ammonia ($\text{Cyt} + \text{H}_2\text{O} \rightarrow \text{Ura} + \text{NH}_3$) is found to be exothermic/exergonic at all levels of theory. The free energy of the deamination reaction is $-34.2 \text{ kJ mol}^{-1}$ at the G3MP2 level of theory. The relative stability of $\text{Cyt} + \text{OH}^-$ versus $\text{Ura}^- + \text{NH}_3$ ($\Delta H = -223.3 \text{ kJ mol}^{-1}$ at G3MP2) reflects the large stabilization of the negative charge for Ura^- versus OH^- . It should be noted that in this case the inclusion of diffuse functions is essential, changing the relative stability by more than 100 kJ mol^{-1} . The relative stability of $[(\text{Cyt}^- \cdots 2\text{H}_2\text{O})$ versus $(\text{Ura}^- \cdots \text{H}_2\text{O} \cdots \text{NH}_3)]$ is similar to the relative stability of $[(\text{Cyt} \cdots 2\text{H}_2\text{O})$ versus $(\text{Ura} \cdots \text{H}_2\text{O} \cdots \text{NH}_3)]$ differing by no more than 5.1 kJ mol^{-1} at G3MP2 indicating that the cytosine and uracil anions are of similar stabilities. The enthalpy for $(\text{Cyt} + \text{H}_2\text{O} \rightarrow \text{Ura} + \text{NH}_3)$ and $[(\text{Cyt} \cdots 2\text{H}_2\text{O})$ versus $(\text{Ura} \cdots \text{H}_2\text{O} \cdots \text{NH}_3)]$ are -35.5 and $-37.6 \text{ kJ mol}^{-1}$ at G3MP2 level of theory, respectively differing by only 2.1 kJ mol^{-1} . This suggests that the enthalpy for the deamination of hydrogen bonded cytosine is not likely to differ significantly compared to the gas phase value.

5.4 Conclusions

Mechanisms for the deamination reaction of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ and $2\text{H}_2\text{O}/\text{OH}^-$ to produce uracil was investigated using *ab initio* calculations. Optimized geometries of reactants, transition states, intermediates, and products were determined at RHF, MP2, and B3LYP using the 6-31G(d) basis set and at B3LYP/6-31+G(d) levels of theory. Single point energies were also determined at MP2/GTMP2Large and G3MP2 levels of theory. Thermodynamic properties (ΔE , ΔH , and ΔG), activation energies, enthalpies and free energies of activation were calculated for each reaction pathway investigated. Intrinsic reaction coordinate (IRC) analysis was performed to characterize the transition states on the potential energy surface.

This work shows the first detailed study of possible mechanisms for the deamination reaction of cytosine with OH^- including the effect of the presence of one and two water molecules. Seven pathways for the deamination reaction were found. All pathways produce an initial tetrahedral intermediate followed by several conformational changes. The final intermediate for all pathways dissociates to products via a 1–3 proton shift. In all the mechanisms, a series of conformational changes connect the transition states of the two rate-determining steps. Pathway D appears to be a plausible mechanism for the deamination of cytosine and in excellent agreement with the experimentally determined activation energy ($117 \pm 4 \text{ kJ mol}^{-1}$). We found that the activation energy of the deamination of cytosine for pathway

D, the first mechanism reported that can account for the observed experimental activation energy, is $115.3 \text{ kJ mol}^{-1}$ at G3MP2 level of theory.

5. 5 References

- (1) Saenger, W. *Principles of Nucleic Acid Structure*: Springer-Verlag: New York, **1983**.
- (2) Neidle, S. *Oxford Hand Book of Nucleic Acid Structure*; Oxford University Press Inc.: New York, **1999**.
- (3) Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2006**, 110, 8227-8234.
- (4) Peng, W.; Shaw, B. R. *Biochemistry* **1996**, 35, 10172-10181.
- (5) Notari, R. E.; Chin, M. L.; Cardoni, A. *J. Pharm. Sci.* **1970**, 59, 28.
- (6) Yao, L.; Li, Y.; Wu, Y.; Liu, A.; Yan, H. *Biochemistry* **2005**, 44, 5940-5947.
- (7) Shapiro, R.; Servis, R. E.; Wecher, M. *J. Am. Chem. Soc.* **1970**, 92, 422.
- (8) Shapiro, R.; Klein, R. *Biochemistry* **1967**, 6, 3576.
- (9) Brown, D.; Phillips, J. H. *J. Mol. Biol.* **1965**, 11, 663-671.
- (10) Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, 98, 6338.
- (11) Shapiro, R.; Klein, R. *Biochemistry* **1966**, 5, 2358.
- (12) Chen, H.; Shaw, B. R. *Biochemistry* **1994**, 33, 4121-4129.
- (13) Bobek, M.; Cheng, Y. C.; Bloch, A. *J. Medicinal chemistry* **1978**, 21, 597.
- (14) Duncan, B. K.; Miller, J. H. *Nature* **1980**, 287, 560-561.
- (15) Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. *Biochemistry* **1990**, 29, 2532-2537.
- (16) Lindahl, T.; Nyberg, B. *Biochemistry* **1974**, 13, 3405-3410.

- (17) Sponer, J. E.; Miguel, P. J.; Rodriguez-Santiago, L.; Erxleben, A.; Krumm, M.; Sodupe, M.; Sponer, J.; Lippert, B. *Angew. Chem. Int. Ed.* **2004**, 43, 5396-5399.
- (18) Rayat, S.; Qian, M.; Glaser, R. *Chem. Res. Toxicol.* **2005**, 18, 1211-1218.
- (19) Matsubara, T.; Ishikura, M.; Aida, M. *J. Chem. Inf. Model.* **2006**, 46, 1276-1285.
- (20) Flinn, C.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2003**, 107, 11174-11181.
- (21) Fogarasi, G.; Szalay, P. G. *Chem. Phys. Lett.* **2002**, 356, 383.
- (22) Frisch, M. J. et al. *Gaussian 03*, Revision B.05, Gaussian, Inc.:Pittsburgh PA, **2003**.
- (23) Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A. *Can. J. Chem.* **2005**, 83, 2082-2090.
- (24) Snider, M. J.; Reinhardt, L.; Wolfenden, R.; Cleland, W. W. *Biochemistry* **2002**, 41, 415-421.
- (25) Florián, J.; Baumruk, V.; Leszczyński, J. *J. Phys. Chem.* **1996**, 100, 5578-5589.
- (26) Hunter, K. C.; Wetmore, S. D. *Chem. Phys. Lett.* **2006**, 422, 500-506.
- (27) Hunter, K. C.; Rutledge, L. R.; Wetmore, S. D. *J. Phys. Chem. A* **2005**, 109, 9554-9562.
- (28) Tureček, F.; Yao, C. *J. Phys. Chem. A* **2003**, 107, 9221-9231.
- (29) Zhang, H.; Liang, Q.; Xia, Y.; Zhao, M.; Ji, Y.; Song, C.; Liu, X.; Zhang, B. *Int. J. Quantum Chem.* **2007**, 107, 240-246.

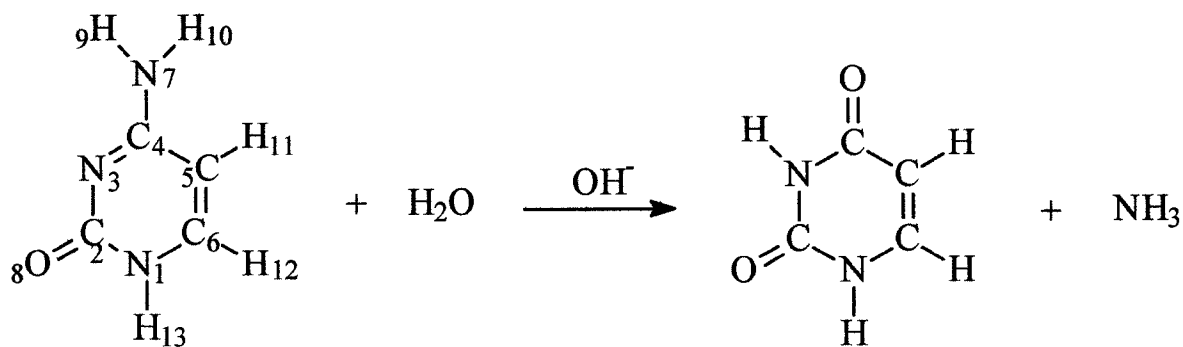


Figure 5.1: Deamination of cytosine with H₂O/OH⁻ and the atom labeling in cytosine.

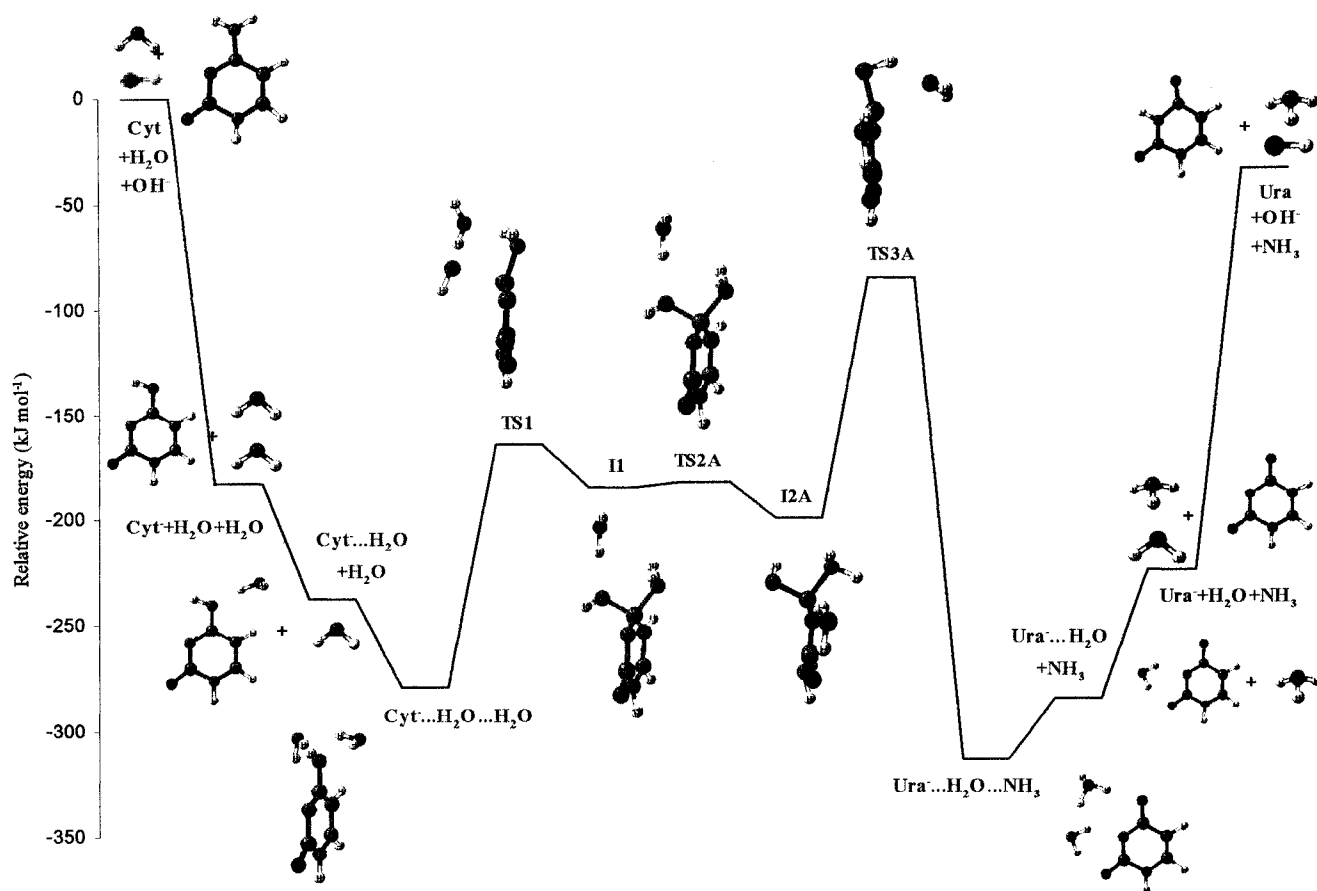


Figure 5.2: Reaction pathway A with the optimized structures for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory).

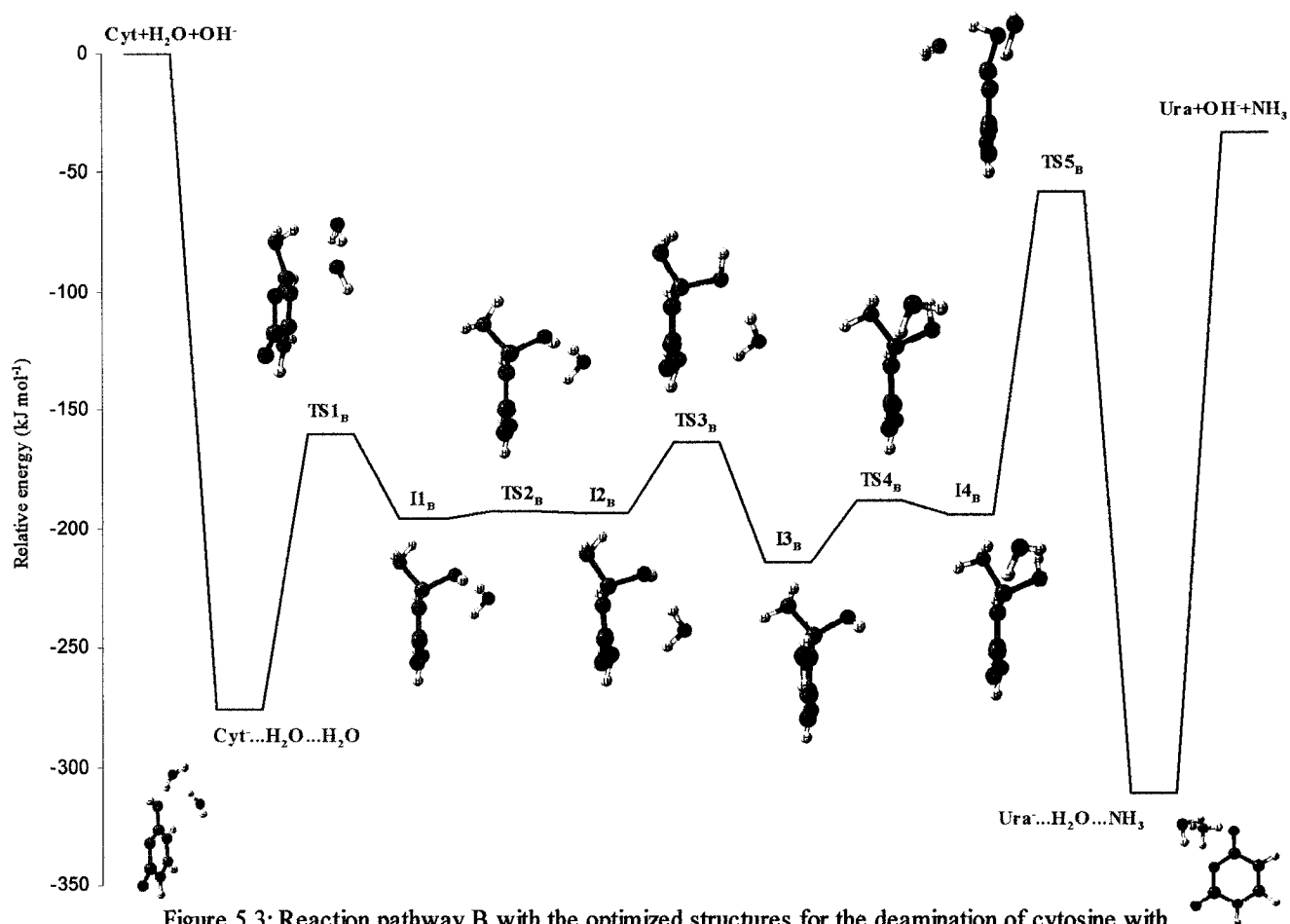


Figure 5.3: Reaction pathway B with the optimized structures for the deamination of cytosine with $\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory).

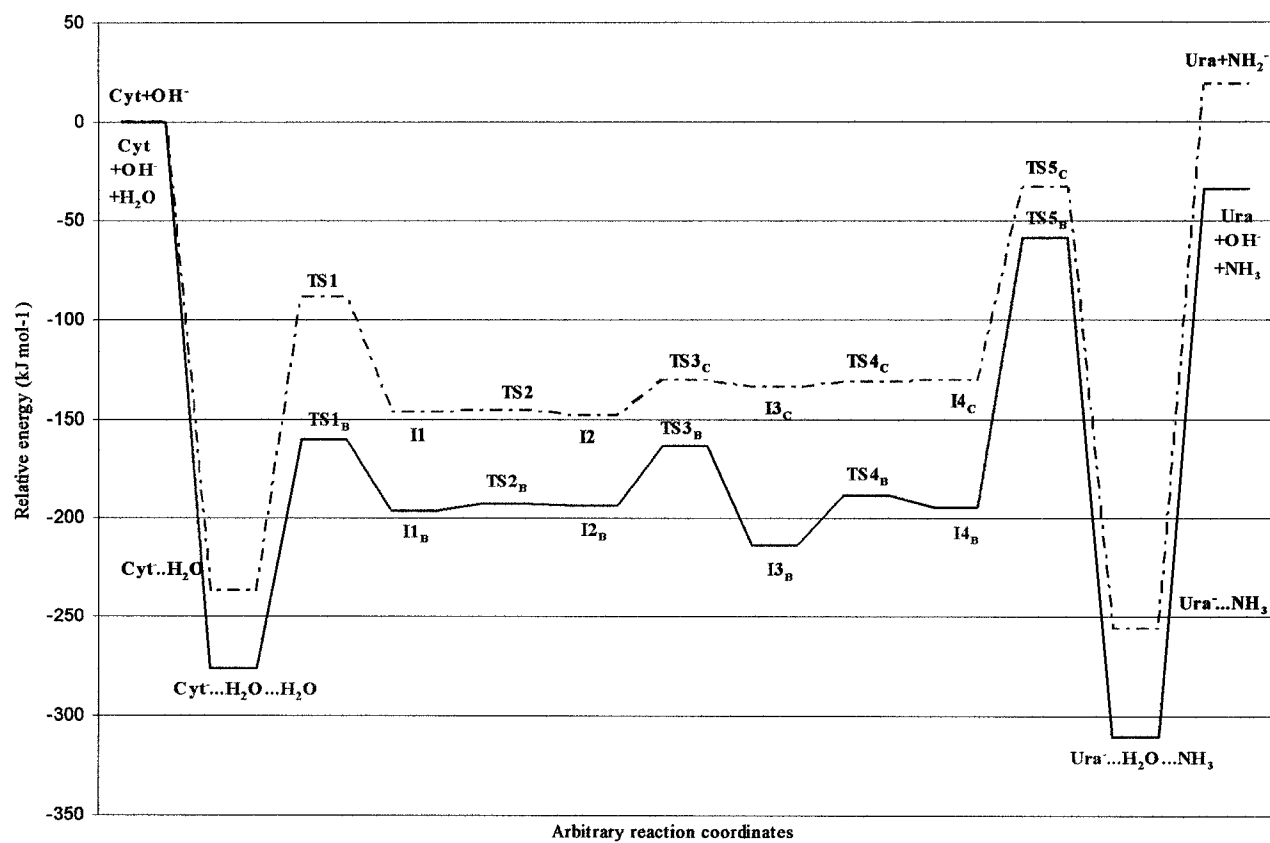


Figure 5.4: Comparison of reaction pathways for the deamination of cytosine with OH⁻ (dotted line) [3] and reaction pathway B for the deamination of cytosine with H₂O/OH⁻ at G3MP2 level of theory.

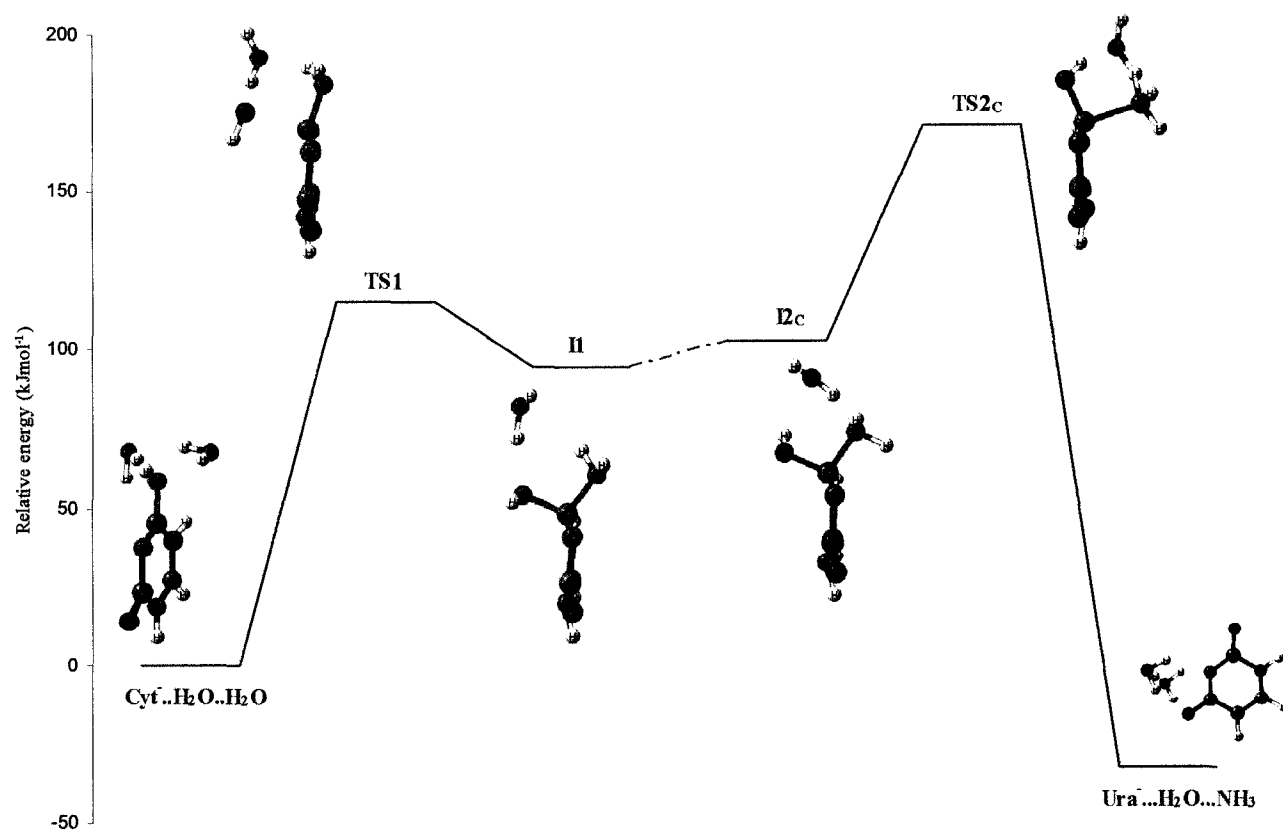


Figure 5.5: Reaction pathway C with the optimized structures for the deamination of cytosine with H₂O/OH⁻ (Relative energies at G3MP2 level of theory).

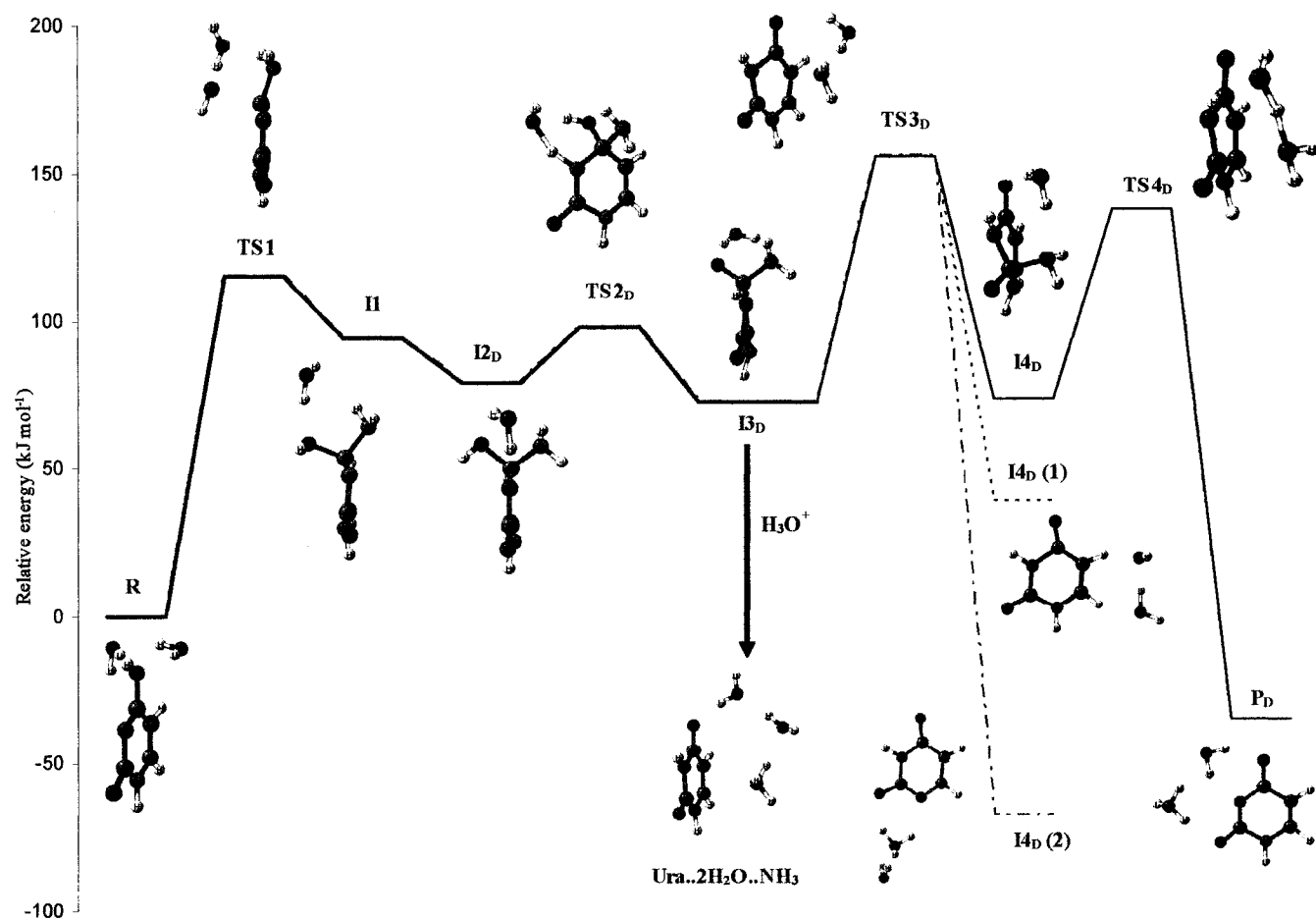


Figure 5.6: Reaction pathway D with the optimized structures for the deamination of cytosine with H₂O/OH⁻ (Relative energies at G3MP2 level of theory) (bold line). I4D(1) optimized structure at B3LYP/6-31G(d) and I4D(2) at B3LYP/6-31+G(d) and HF/6-31G(d)

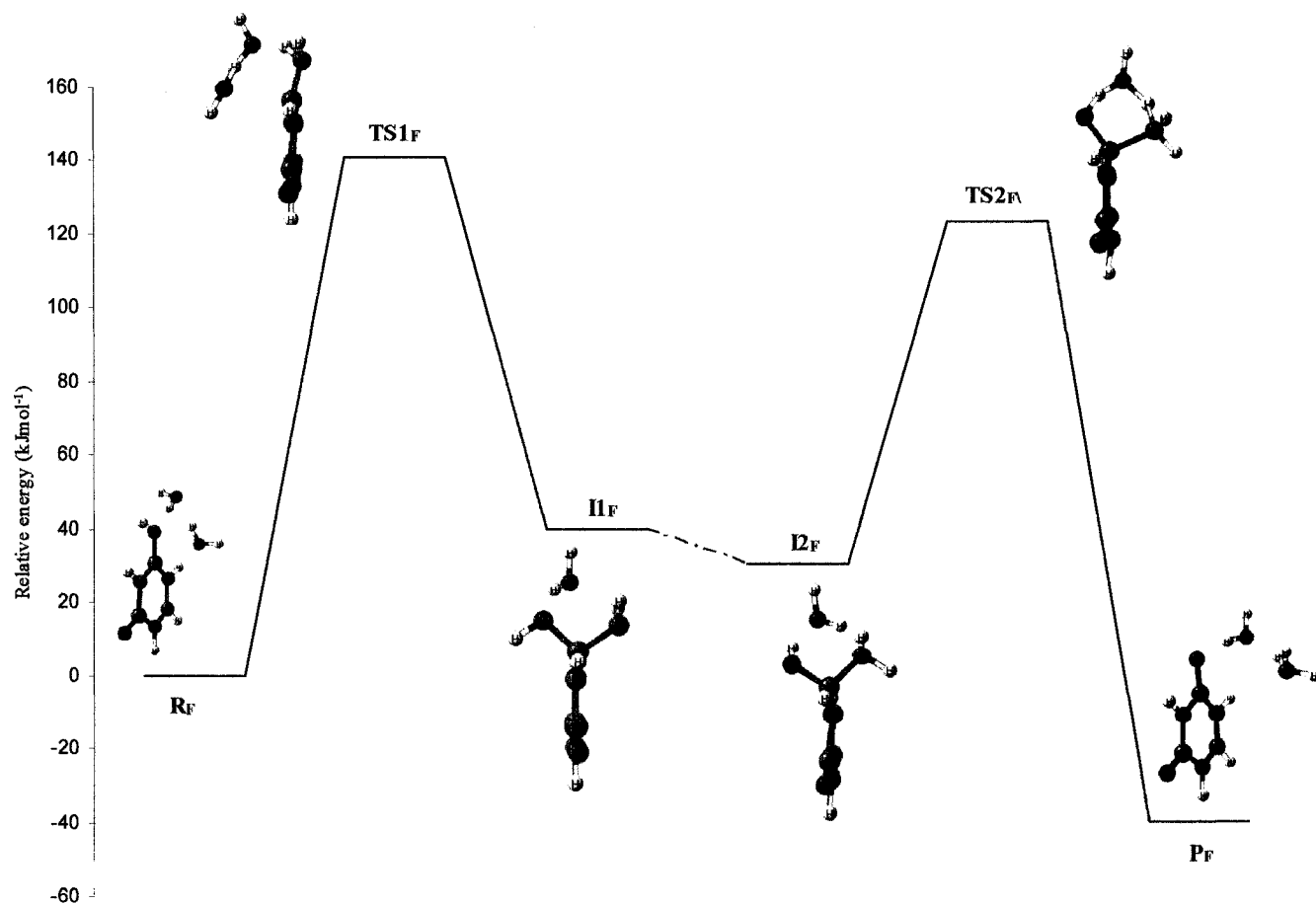


Figure 5.7: Reaction pathway F with the optimized structures for the deamination of the amino-oxo tautomer of cytosine with 2H₂O (Relative energies at G3MP2 level of theory).

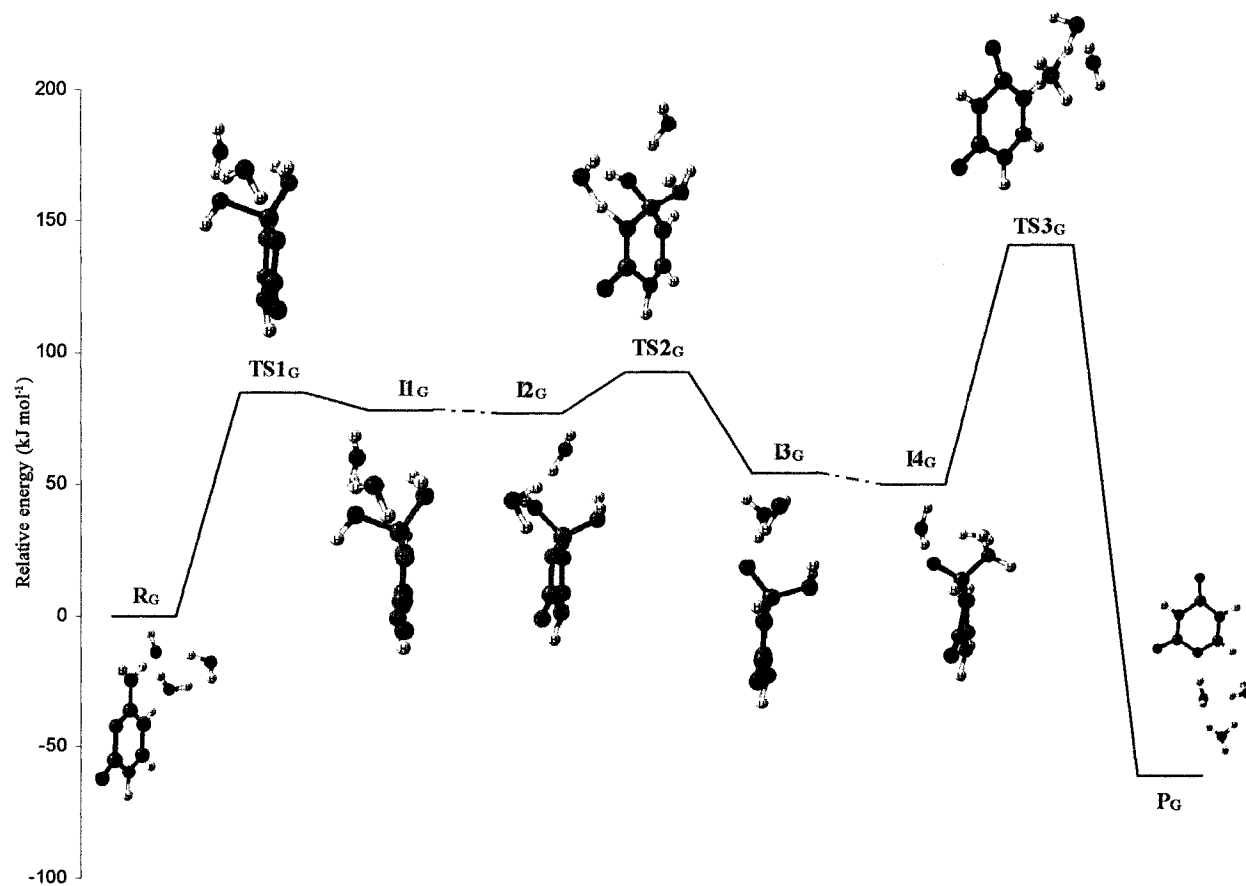


Figure 5.8: Reaction pathway G with the optimized structures for the deamination of cytosine with 2H₂O/OH⁻ (Relative energies at G3MP2 level of theory).

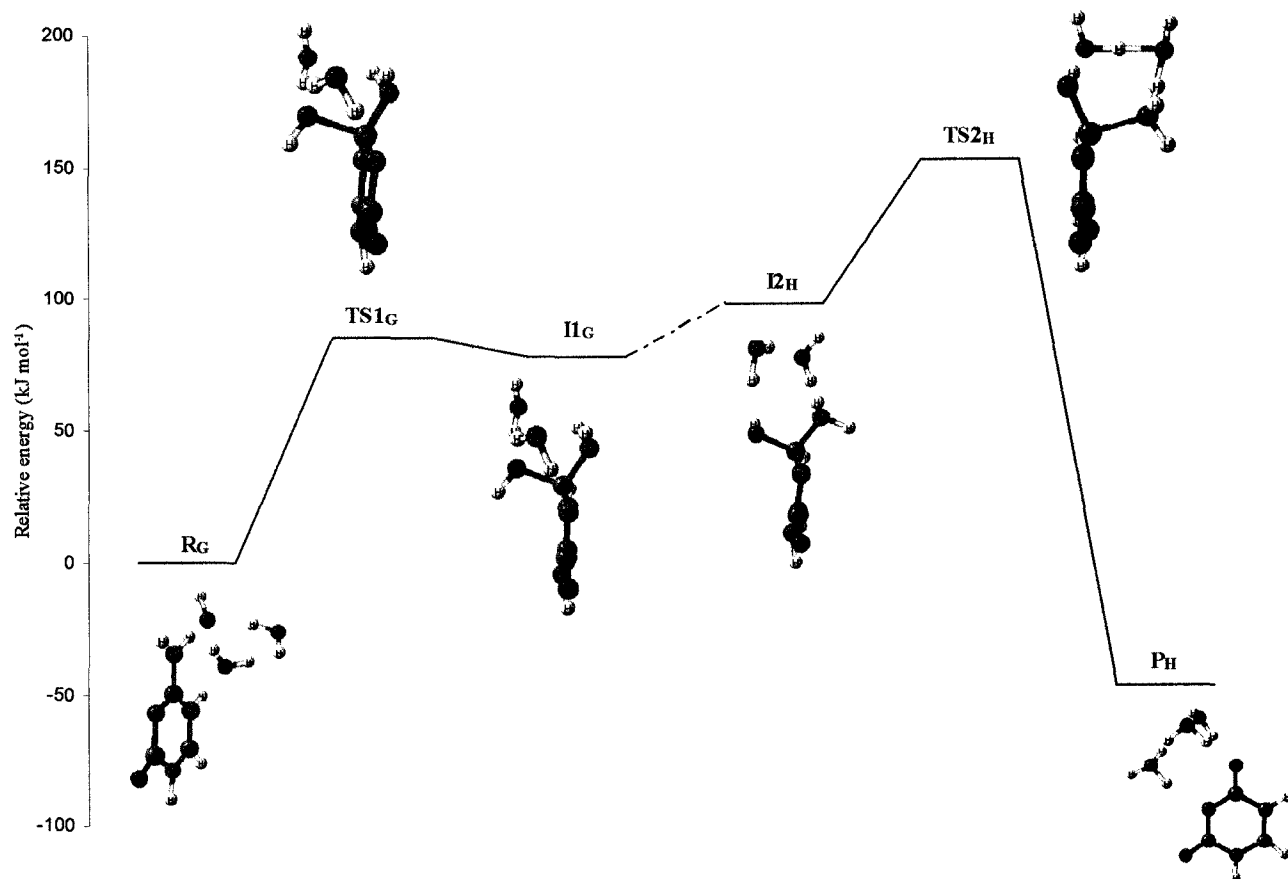


Figure 5.9: Reaction pathway H with the optimized structures for the deamination of cytosine with $2\text{H}_2\text{O}/\text{OH}^-$ (Relative energies at G3MP2 level of theory).

TABLE 5.1: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H₂O/OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway A).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{act}, TS1$	159.3	137.1	130.7	118.4	115.3 (116.5) ^a
$\Delta H^\ddagger, TS1$	154.2	129.8	124.6	114.7	112.1
$\Delta G^\ddagger, TS1$	163.8	139.2	133.4	122.9	122.0
$\Delta E_{act}, TS2_A$	4.9	8.2	7.5	4.6	2.5 (4.4)
$\Delta H^\ddagger, TS2_A$	2.7	5.6	4.4	2.4	0.9
$\Delta G^\ddagger, TS2_A$	3.9	5.1	4.6	7.7	2.5
$\Delta E_{act}, TS3_A$	154.8	141.7	107.6	103.7	114.0 (125.8)
$\Delta H^\ddagger, TS3_A$	140.4	122.6	92.9	90.9	115.1
$\Delta G^\ddagger, TS3_A$	139.2	118.1	89.4	87.8	113.8

^a The values in bracket are for MP2/GTMP2Large.

TABLE 5.2: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H₂O/OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway B).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{act}, TS1_B$	153.4	132.6	127.3	116.8	115.7 (120.2) ^a
$\Delta H^\ddagger, TS1_B$	148.5	129.3	122.7	114.0	113.3
$\Delta G^\ddagger, TS1_B$	157.3	144.4	135.8	123.3	122.4
$\Delta E_{act}, TS2_B$	2.7	8.5	2.3	2.7	3.5 (4.2)
$\Delta H^\ddagger, TS2_B$	0.5	5.4	0.2	0.7	2.0
$\Delta G^\ddagger, TS2_B$	3.3	6.5	3.4	3.9	5.0
$\Delta E_{act}, TS3_B$	42.7	45.4	41.4	35.1	29.9 (34.1)
$\Delta H^\ddagger, TS3_B$	38.2	41.2	36.8	30.4	29.0
$\Delta G^\ddagger, TS3_B$	38.8	42.1	37.5	30.8	29.9
$\Delta E_{act}, TS4_B$	32.8	26.5	23.5	15.0	26.0 (26.3)
$\Delta H^\ddagger, TS4_B$	29.9	23.9	20.7	11.8	24.2
$\Delta G^\ddagger, TS4_B$	33.6	29.7	26.3	15.1	28.3
$\Delta E_{act}, TS5_B$	155.9	144.7	112.2	100.7	115.2 (128.3)
$\Delta H^\ddagger, TS5_B$	140.6	124.6	95.9	88.1	113.5
$\Delta G^\ddagger, TS5_B$	134.0	115.3	88.5	84.7	70.5

^a The values in bracket are for MP2/GTMP2Large.

TABLE 5.3: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H₂O/OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway C).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{act,TS1}$	159.3	137.1	130.7	118.4	115.3 (116.5) ^a
$\Delta H^\ddagger, TS1$	154.2	129.8	124.6	114.7	112.1
$\Delta G^\ddagger, TS1$	163.8	139.2	133.4	122.9	122.0
$\Delta E_{act,TS2_C}$	105.1	73.0	52.2	50.2	68.3 (65.3)
$\Delta H^\ddagger, TS2_C$	96.6	59.7	36.6	41.4	65.2
$\Delta G^\ddagger, TS2_C$	106.2	66.3	42.1	51.7	68.3

^a The values in bracket are for MP2/GTMP2Large.

TABLE 5.4: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with H₂O/OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway D).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{act}, TS1$	159.3	137.1	130.7	118.4	115.3 (116.5)^a
$\Delta H^\ddagger, TS1$	154.2	129.8	124.6	114.7	112.1
$\Delta G^\ddagger, TS1$	163.8	139.2	133.4	122.9	122.0
$\Delta E_{act}, TS2_D$	56.4	37.9	27.0	21.6	19.2 (25.3)
$\Delta H^\ddagger, TS2_D$	41.9	23.8	12.3	8.0	16.6
$\Delta G^\ddagger, TS2_D$	47.7	28.6	16.5	16.1	22.8
$\Delta E_{act}, TS3_D$	115.3	120.9	95.2	69.4	82.4 (93.6)
$\Delta H^\ddagger, TS3_D$	103.3	108.3	76.7	58.0	83.9
$\Delta G^\ddagger, TS3_D$	101.1	107.3	68.4	59.4	81.5
$\Delta E_{act}, TS4_D$	77.1	87.7	55.0	73.4 ^b	63.6 (78.7)
$\Delta H^\ddagger, TS4_D$	59.0	73.4	47.0	51.6	64.4
$\Delta G^\ddagger, TS4_D$	56.9	78.8	59.4	58.0	62.3

^a The values in bracket are for MP2/GTMP2Large. Bold numbers are indicating the rate-determining step for this pathway.

^b TS4_D did not converge to a first-order saddle point at B3LYP/6-31+G(d) but converged to the product, the value reported is for B3LYP/6-31+G(d)// HF/6-31G(d).

TABLE 5.5: Activation energies, enthalpies of activation and free energies of activation for the deamination of the amino-oxo tautomer of cytosine with H₂O/OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway F).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{act}, TS1_F$	200.4	155.9	140.4	129.6	140.9 (141.8) ^a
$\Delta H^\ddagger, TS1_F$	184.9	140.9	125.0	116.3	135.1
$\Delta G^\ddagger, TS1_F$	203.0	159.4	140.1	129.5	153.8
$\Delta E_{act}, TS2_F$	151.4	93.8	80.6	81.6	92.6 (91.1)
$\Delta H^\ddagger, TS2_F$	135.2	74.4	61.4	63.4	88.2
$\Delta G^\ddagger, TS2_F$	145.6	80.8	68.0	70.7	99.2

^a The values in bracket are for MP2/GTMP2Large.

TABLE 5.6: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with $2\text{H}_2\text{O}/\text{OH}^-$ (in kJ mol^{-1}) at 298.15 K (Pathway G).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{\text{act}}, \text{TS1}_G$	105.9	76.1	96.6	110.9	84.8 (86.5) ^a
$\Delta H^\ddagger, \text{TS1}_G$	106.2	79.3	99.8	112.3	82.0
$\Delta G^\ddagger, \text{TS1}_G$	117.5	91.1	112.6	123.9	93.5
$\Delta E_{\text{act}}, \text{TS2}_G$	47.9	22.8	16.4	16.7	16.2 (18.9)
$\Delta H^\ddagger, \text{TS2}_G$	35.2	9.9	4.2	4.7	12.4
$\Delta G^\ddagger, \text{TS2}_G$	45.8	17.4	10.6	12.5	23.4
$\Delta E_{\text{act}}, \text{TS3}_G$	133.7	106.1	83.0	75.5	90.3 (94.5)
$\Delta H^\ddagger, \text{TS3}_G$	119.4	86.7	63.8	61.4	90.2
$\Delta G^\ddagger, \text{TS3}_G$	118.6	83.6	58.1	54.8	89.5

^a The values in bracket are for MP2/GTMP2Large.

TABLE 5.7: Activation energies, enthalpies of activation and free energies of activation for the deamination of cytosine with 2H₂O/OH⁻ (in kJ mol⁻¹) at 298.15 K (Pathway H).

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
$\Delta E_{act, TS1_H}$	105.9	76.1	96.6	110.9	84.8 (86.5) ^a
$\Delta H^\ddagger, TS1_H$	106.2	79.3	99.8	112.3	82.0
$\Delta G^\ddagger, TS1_H$	117.5	91.1	112.6	123.9	93.5
$\Delta E_{act, TS2_H}$	90.3	57.1	35.2	57.6	55.3 (50.6)
$\Delta H^\ddagger, TS2_H$	80.5	41.3	21.4	46.9	51.5
$\Delta G^\ddagger, TS2_H$	92.1	53.8	30.5	55.9	63.6

^a The values in bracket are for MP2/GTMP2Large.

TABLE 5.8: Thermodynamic properties and relative stabilities of several species investigated (in kJ mol⁻¹) at 298.15 K.

	HF/ 6-31G(d)	MP2/ 6-31G(d)	B3LYP/ 6-31G(d)	B3LYP/ 6-31+G(d)	G3MP2
(Cyt+H₂O) versus (Ura+ NH₃)					
ΔE	-75.0	-64.8	-70.7	-58.6	-34.1 (-36.7) ^a
ΔH	-71.0	-60.6	-66.7	-54.4	-35.5
ΔG	-70.0	-60.7	-66.1	-53.2	-34.2
(Cyt+OH⁻) versus (Ura⁻+NH₃)					
ΔE	-328.2	-339.6	-347.8	-229.9	-223.0 (-224.9)
ΔH	-325.2	-337.6	-346.0	-229.3	-223.3
ΔG	-329.4	-344.5	-351.1	-233.8	-227.1
(Cyt⁻...2H₂O) versus (Ura⁻...H₂O...NH₃)					
ΔE	-69.0	-53.1	-57.2	-47.9	-34.6 (-29.7)
ΔH	-68.6	-51.1	-55.5	-45.2	-32.6
ΔG	-73.9	-56.8	-60.4	-53.1	-38.0
(Cyt...2H₂O)^b versus (Ura...H₂O...NH₃)					
ΔE	-58.1	-54.6	-43.0	-46.3	-39.7 (-36.6)
ΔH	-58.5	-51.5	-41.6	-44.9	-37.6
ΔG	-66.8	-52.0	-51.2	-50.8	-46.0

^a The values in bracket are for MP2/GTMP2Large.

^b Cytosine tautomer with two water molecules.

CHAPTER 6

CONCLUDING REMARKS

We have investigated the deamination reaction of cytosine and the decomposition reaction of formamidine by performing high level *ab initio* calculations, Gaussian-n theories, and DFT calculations. Optimized geometries of reactants, transition states, intermediates, and products were determined at HF, MP2, and B3LYP using the 6-31G(d) and 6-31+G(d) basis sets. Single point energies were also calculated at MP2/GTMP2Large and Gaussian-n theories. Thermodynamics properties (ΔE , ΔH , and ΔG), activation energies, enthalpies and free energies of activation were calculated for each reaction pathway investigated. Intrinsic reaction coordinate (IRC) analysis was performed, using *ab initio* and DFT methods, to characterize all the optimized transition states on the potential energy surface (PES). The PCM solvent models were used to study the effect of solvent on the decomposition reaction of formamidine.

The computational study of the decomposition of formamidine contributed toward the understanding of the mechanism of proton transfer reactions in hydrogen-bonded systems. The decomposition reaction of formamidine, yielding hydrogen cyanide and ammonia, has been studied first as a simple model for the intramolecular and intermolecular hydrogen rearrangement of cytosine. This reaction has yielded several conclusions:

1. The high degree of stability of formamidine toward decomposition to ammonia and hydrogen cyanide is confirmed by this investigation.
2. The gas phase decomposition of formamidine predicted a high activation energy of 259 kJ mol^{-1} at the G3 level of theory. The high activation energies explain why rather extreme conditions are necessary to achieve this reaction experimentally.
3. We found that adding one water molecule catalyses the reaction by forming a cyclic hydrogen bonded transition state, reducing the barrier to 169 kJ mol^{-1} at the G3 level. Addition of a second water molecule, further reduces the barrier to 151 kJ mol^{-1} at the G3 level.
4. The barrier for the water-mediated 1-3 proton shift for the decomposition of formamidine dropped by approximately 90 kJ mol^{-1} at Gaussian-n theories.
5. The decomposition reaction of formamidine with separated products is more exergonic than the formation of the HCN/NH₃ complex due to the increase in the entropy for the formation of the separated products.
6. A two water-mediated transition state has not been found despite a thorough investigation.
7. Concerning the role of the solvent, we should point out that it is dangerous in the case of

mechanistic studies to assume that the reaction mechanisms in the gas phase (single-point energy calculations using gas phase optimized structures) and in solution are the same. The decomposition reaction of formamidine in solution involves a two-step mechanism for both the unimolecular and the water-catalyzed decomposition compared to one step in the gas phase. A zwitterionic intermediate existed for this reaction in solution which is stabilized in solution with regard to the gas phase.

8. The PCM solvent model predicts a significant lowering of the free energy of activation: 16-18 kJ mol⁻¹ for the unimolecular reaction; and 21-42 kJ mol⁻¹ for the water-mediated reaction in aqueous solution relative to the gas phase.
9. We found that the activation energies and the heats of reaction calculated using Gaussian-n theories (G1, G2, G2MP2, G3, G3MP2, G3B3, and G3MP2B3) all agreed to within 10 kJ mol⁻¹. On the basis of these results, we chose the G3MP2 level of theory which is computationally the least expensive method to give reliable energetics.

The mechanism for the deamination reaction of cytosine with H₂O and OH⁻ to produce uracil was investigated using *ab initio* and DFT calculations. Single-point energies were also determined at B3LYP/6-31+G(d), MP2/GTMP2Large and the G3MP2 levels of theory. All pathways, in the cytosine deamination, produce an initial tetrahedral intermediate followed by several conformational changes. The final intermediate for all pathways dissociates to products via a 1-3 proton shift from the hydroxyl group to the exocyclic nitrogen atom (amino group). Two pathways for the deamination reaction of cytosine with H₂O were found. The activation

energy for the rate-determining steps, the formation of a tetrahedral intermediate for pathway A, and the formation of the uracil tautomer for pathway B are 221 and 260 kJ mol⁻¹ at G3MP2, respectively. The deamination reaction by either pathway is therefore unlikely because of the high barriers involved. Pathways C and D for deamination with OH⁻ were also found. Both pathways produce an initial tetrahedral intermediate by adding H₂O to deprotonated cytosine which then undergoes three conformational changes. Deamination with OH⁻ through pathway C resulted in the lowest activation energy, 203 kJ mol⁻¹ at G3MP2. However this value is still high compared to the experimental value (117 ± 4 kJ mol⁻¹).

The deamination reaction of cytosine with H₂O/OH⁻ and 2H₂O/OH⁻, in which the water molecules act as a solvent and a catalyst, was investigated using *ab initio* and DFT calculations. Single-point energies were also determined at the MP2/GTMP2Large and the G3MP2 levels of theory. The mechanism for the reaction with H₂O/OH⁻ and 2H₂O/OH⁻ closely follows the mechanism for the deamination with H₂O and the deamination with OH⁻, particularly in relation to the rate-determining steps. A nucleophilic attack by H₂O forms a tetrahedral intermediate followed by a conformational change and a second rate-determining step, which is a 1–3 proton shift. Seven pathways for this reaction were found. We also considered the possibility of the deamination of a tautomer of cytosine with 2H₂O being protonated. The activation energy for the rate-determining step, the formation of tetrahedral intermediate for pathway D, the only pathway that can lead to uracil is 115 and 118 kJ mol⁻¹ at G3MP2 and B3LYP/6-31+G(d) levels of theory, respectively. This result is in excellent agreement with the experimental value.

For the deamination reaction of cytosine, we can make several remarks:

1. This work shows, for the first time, a plausible mechanism for the deamination of cytosine and accounts for the observed experimental activation energy.
2. The complete reaction pathway for all the proposed mechanisms have been verified using IRC analyses of all transition states.
3. We found that there were no changes in the potential energy surface, except in pathway D, for these reactions when using both density functional theory and wavefunction based methods.
4. Of all the pathways investigated, intermediate $I3_D$ and $I3_G$, are the only intermediate that can be protonated and lead to neutral uracil.
5. The water molecule can play an important role in cytosine deamination in DNA.
6. We found that the barrier for the water-mediated 1–3 proton shift for cytosine deamination is reduced by 46 kJ mol^{-1} at G3MP2 and 54 kJ mol^{-1} at B3LYP/6-31+G(d). We also found that the addition of the second water molecule reduces the barriers for both rate-determining steps by 31 kJ mol^{-1} .
7. For the reaction involving separate reactants, the addition of diffuse functions is shown here to be essential, changing the relative stability by more than 100 kJ mol^{-1} .
8. On the basis of our data in this thesis, we found that the B3LYP/6-31+G(d) results are in good agreement with the G3MP2 results differing by no more than 18 kJ mol^{-1} . Thus, the computationally less expensive method, B3LYP/6-31+G(d), will be a good choice to study large molecules for similar systems, such as investigating the deamination of cytidine and

other large molecules.

9. The MP2/GTMP2Large results for both the thermochemistry and barriers are in excellent agreement with the G3MP2 values, differing by no more than 13 kJ mol⁻¹.
10. The enthalpy for the deamination of hydrogen bonded cytosine is not likely to differ significantly compared to the gas phase value.

Possible future work includes:

1. On the basis of the insights gained from the thorough study of the deamination reaction of cytosine, we will proceed to study the effect of different catalysts on the deamination reaction of cytosine.
2. We now can study the deamination of cytidine with B3LYP/6-31+G(d) as a reliable method.
3. We may study the effect of solvent on the deamination of cytosine and cytidine by PCM solvent models.
4. We may study the deamination of other nucleic acid bases, such as guanine.
5. We will investigate systems related to DNA damage and repair.

Appendix A

This table shows the number of basis functions and some other data for the cytosine (Cyt) and cytosine anion (Cyt⁻) at HF/6-31G(d) and HF/6-31+G(d) levels of theory.

	Cyt	Cyt	Cyt ⁻	Cyt ⁻
	HF/6-31G(d)	HF/6-31+G(d)	HF/6-31G(d)	HF/31+G(d)
Basis functions	130	162	128	160
Primitive Gaussians	244	276	240	272
Point group	C ₁	C ₁	C ₁	C ₁
One-electron integrals ^a	8515	13203	8256	12880
Two-electron integrals ^b	22,718,751	65,624,398	17,998,911	55,112,404
Deg. of freedom	33	33	30	30
Chemical Formula	C ₄ H ₅ N ₃ O	C ₄ H ₅ N ₃ O	C ₄ H ₄ N ₃ O ⁻	C ₄ H ₄ N ₃ O ⁻

^a One-electron integrals calculated using $[N(N+1)/2]$, where N is the number of the basis functions.

^b These numbers represent the actual number of two-electron integrals from MUNgauss quantum chemical packages.

Appendix B

This table shows a comparison of some properties between different levels of theory:

Theory	Variational	Size-Extensive	Cost ^a
HF	Yes	Yes	$O(N^2)^b$
DFT ^c	Yes	Yes	$O(N^3)$
MPn	No	Yes	$O(N^{n+3})$
Full CI	Yes	Yes	$O(N_{\text{det}} N^4)^d$
Truncated CI	Yes	No	$O(N^{2m+2})^e$
QCISD(T)	Yes	Yes	$O(N^7)$
CCSD	No	Yes	$O(N^6)$
CCSDT	No	Yes	$O(N^8)$

^a Scaling refers to how time increases with number of basis functions (N).

^b HF formally scales as $O(N^4)$, practically as $O(N^2)$.

^c The exact DFT is variational but this is not true once approximation for E_{XC} are adopted.

^d The full CI scales as $O(N_{\text{det}} N^4)^7$, where N_{det} is the number of determinants.

^e Where m is the excitation level (e. g., 2 for CISD), thus, CISD scales as $O(N^6)$ and CISDTQ scales as $O(N^{10})$.

⁷ Knowles, P. J; Handy, N. C. *Chem Phys. Lett.* **1984**, 111, 315.

Appendix C

This table shows the energy (in Hartree) and the job cpu time for the Gaussian-n theories for two examples from the decomposition reaction of formamidine with one and two water molecules (Form/H₂O and Form/2H₂O):

	TS1	Job cpu time	TS1	Job cpu time
	Form/H ₂ O	(M.S)	Form/2H ₂ O	(M.S)
RHF	-225.000100	1.25	-301.037824	5.27
MP2	-225.666104	8.37	-301.896416	28.01
G1	-226.020546	134.13	-302.371669	791.30
G2	-226.025676	28.05	-302.380360	64.00
G2MP2	-226.019717	28.05	-302.372124	64.00
G3	-226.214371	40.55	-302.619122	100.34
G3B3	-226.220500	37.27	-302.626708	88.48
G3MP2	-226.059284	22.42	-302.422985	53.23
G3MP2B3	-226.070435	22.39	-302.436983	50.52

(M.S) stands for (Minutes . Seconds)

Appendix D

Manual G1/G2 level of theory calculations for the first transition state of the decomposition reaction of formamidine.

Step No. (N)		Energy (Hartree)
1	MP2/6-311G(d,p)	-149.617397
2	MP2/6-311+G(d,p)	-149.627306
3	MP2/6-311G(2df,p)	-149.692966
4	MP2/6-311+G(2df,p)	-149.702860
5	MP2/6-311+G(3df,2p)	-149.716826
6	MP4/6-311G(d,p)	-149.660745
7	MP4/6-311+G(d,p)	-149.670641
8	MP4/6-311G(2df,p)	-149.740080
9	QCISD(T)/6-311G(d,p)	-149.660226
10	E (+)	-0.009896
11	E (2df)	-0.079335
12	E (QCI)	0.000519
13	E (+,2df)	0.000015
14	E (3df,2p)	-0.013966
15	E (HLC)	-0.055260
16	E (ZPE)	0.056056
17	E(thermal)[R+V+T]	0.059838
18	E (electronic energy)	-149.804716
19	G1(Energy)	-149.744878
20	E (G1) (0 K)	-149.748660
21	E (G2) (0 K)	-149.752351

where,

$$E (+) = N7 - N6$$

$$E(2df) = N8 - N6$$

$$E(QCI) = N9 - N6$$

$$E(+,2df) = (N4 - N2) - (N3 - N1)$$

$$(3df,2p) = N5 - N4$$

$$E(G1 \text{ Energy}) = N17 + N18$$

$$E (\text{electronic energy}) = N9 + N10 + N11 + N15$$

$$E(G1)(0 \text{ K}) = N9 + N10 + N11 + N15 + N16$$

$$E(G2)(0 \text{ K}) = (N5 - N3 - N2 + N1) + N20 + (0.00114 \times 9)$$

Note: for more detail about these calculations, see references [31–33].

For comparison, see the output for the direct optimization (using G3MP2 command direct as implemented in Gaussian03) for the first transition state from the output of Gaussian03:

Temperature =	298.150000	Pressure =	1.000000
E(ZPE) =	0.056056 E	(Thermal) =	0.059838
E(QCISD(T)) =	-149.660226	E(Empiric) =	-0.055260
DE(Plus) =	-0.009896	DE(2DF) =	-0.079335
G1(0 K) =	-149.748660	G1 Energy =	-149.744878
G1 Enthalpy =	-149.743934	G1 Free Energy =	-149.773143
E(Delta-G2) =	-0.013951 E	(G2-Empiric) =	0.010260
G2(0 K) =	-149.752351	G2 Energy =	-149.748569
G2 Enthalpy =	-149.747625	G2 Free Energy =	-149.776834
DE(MP2) =	-0.099428		
G2MP2(0 K) =	-149.748598	G2MP2 Energy =	-149.744816
G2MP2 Enthalpy =	-149.743872	G2MP2 Free Energy =	-149.773082

Manual G3MP2 level of theory calculations for the first transition state of the deamination reaction of cytosine with H₂O/OH⁻.

Step No. (N)		Energy (Hartree)
1	HF/6-31G(d)	-544.03462
2	ZPE	0.14172
3	ZPE *(0.8929)	0.12654
4	Thermal correction for E	0.15224
5	MP2(FC)/6-31G(d)	-545.57988
6	MP2(FULL)/6-31G(d)	-545.61901
7	MP2/GTMP2Large	-546.24853
8	QCISD(T)/6-31G(d)	-545.67814
9	ΔE (MP2)	-0.66865
10	MO's	29.00000
11	A (Hartree)	0.00928
12	E (HLC)	-0.26909
13	E (G3MP2) (0 K)	-546.48933

where,

$$\Delta E \text{ (MP2)} = N7 - N5$$

$$E(\text{HLC}) = -(N10 * N11)$$

$$E(\text{G3MP2}) = N3 + N8 + N9 + N12$$

Note: see equations (51 – 54) for these calculations in section (2.10).

Appendix E

Calculating Free Energies of Reaction or Activation Using Polarizable Continuum Models (PCM) Calculations :

Input keywords :

RHF/6-31G(d) SCF = (Tight) OPT FREQ SCRF = (PCM,Read)

Z-matrix (as usual)

:

RADII = UFF

There are different types of cavities that may be used: PAULING, KLAMT, UA0, UFF....etc.

Notes:

(1) Input file for the single point calculation using the gas phase optimized structures is

RHF/6-31G(d) SCF = (Tight) SCRF = (PCM,Read)

(2) The PCM calculations can be performed using different solvents. The solvent may be written in the input stream in different ways depending on SCRF methods. The default solvent for Gaussian03 is water ($\epsilon = 78.39$). So, if we want to do calculations by using benzene, then the command will be (Solvent = Benzene) or (Solvent = C₆H₆).

Output (from Gaussian03):

This data is given after the optimized structure summary.

SCF Done: **E(RHF) = -225.118233024** A.U. after 1 cycles
Convrg = 0.1668D-08 -V/T = 2.0022
S**2 = 0.0000

Variational PCM results

$\langle \psi(f) | H | \psi(f) \rangle$ (a.u.) = -225.088787

$\langle \psi(f) | H + V(f)/2 | \psi(f) \rangle$ (a.u.) = -225.118233

Total free energy in solution:

with all non electrostatic terms (a.u.) = **-225.109030**

(Polarized solute)-Solvent (kcal/mol) = -18.48

Cavitation energy (kcal/mol) = 14.96

Dispersion energy (kcal/mol) = -10.19

Repulsion energy (kcal/mol) = 1.00

Total non electrostatic (kcal/mol) = 5.77

Also from the thermodynamic results after the frequencies , we get :

Zero-point correction = 0.087836 (Hartree/Particle)

Thermal correction to Energy = 0.095370

Thermal correction to Enthalpy = 0.096314

Thermal correction to Gibbs Free Energy = 0.055211

Sum of electronic and zero-point Energies = -225.030397

Sum of electronic and thermal Energies = -225.022863

Sum of electronic and thermal Enthalpies = -225.021919

Sum of electronic and **thermal Free Energies = -225.063022**

Note:

This example taken from the output of the reactant structure for the decomposition reaction of formamidine with one water molecule using PCM model. Coloured values are needed to calculate the free energy of solution.

See the tables in the next pages for calculation:

E(RHF) = -225.118233024

represents in the table: E_{soln}

free energy in solution = -225.109030

represents in the table: G_{soln}

Thermal correction to Gibbs Free Energy = 0.055211

represents in the table: $\Delta\Delta G_{\text{soln}}$
(thermal)

thermal Free Energies = -225.063022

represents in the table: G_{soln}
(thermal)

An example for calculation the free energy of reaction and activation using PCM model at HF/6-31G(d) level of theory for the decomposition reaction of formamidine with one water molecule.

	R ^a	TS ^a	P ^a	ΔE_{act}^b	ΔE_{rxn}^b
HF / gas phase calculations					
E_{gas}	-225.093024	-225.000100	-225.090389	244.0	6.9
$\Delta\Delta G_{\text{gas}}$ (thermal)	0.056436	0.057341	0.049797	2.4	-17.4
G_{gas} (total)	-225.036588 ^c	-224.942759 ^c	-225.040592 ^c	246.4 ^d	-10.5 ^e
HF / Klamt (single point calculations)					
E_{soln}	-225.117732	-225.035276	-225.107531	216.5	26.8
G_{soln}	-225.108971	-225.028942	-225.095957	210.1	34.2
G_{soln} (total)	-225.052535 ^f	-224.971601 ^f	-225.046160 ^f	212.5 ^g	16.7 ^h
HF / Klamt (for optimized structures)					
E_{soln}	-225.118233	-225.041179	-225.110636	202.3	19.9
$\Delta\Delta G_{\text{soln}}$ (thermal)	0.055211	0.058556	0.048932	8.8	-16.5
G_{soln} (thermal)	-225.063022	-224.982623	-225.061704	211.1	3.5
G_{soln}	-225.109030	-225.034456	-225.098100	195.8	28.7
G_{soln} (total)	-225.053819 ⁱ	-224.975900 ⁱ	-225.049168 ⁱ	204.6 ^j	12.2 ^k

^a Energy in Hartree

^b Energy in kJ mol⁻¹

^c $G_{\text{gas}} = E_{\text{gas}} + \Delta\Delta G_{\text{gas}} \text{ (thermal)}$

^d $\Delta G \text{ (rxn)} = G_{\text{gas}} \text{ (total) [P]} - G_{\text{gas}} \text{ (total) [R]}$

$$^e \Delta G (\text{act}) = G_{\text{gas}} (\text{total}) [\text{TS}] - G_{\text{gas}} (\text{total}) [\text{R}]$$

$$^f G_{\text{soln}} (\text{total}) = G_{\text{soln}} + \Delta\Delta G_{\text{gas}} (\text{thermal})$$

$$^g \Delta G (\text{rxn}) = G_{\text{soln}} (\text{total}) [\text{P}] - G_{\text{soln}} (\text{total}) [\text{R}]$$

$$^h \Delta G (\text{act}) = G_{\text{soln}} (\text{total}) [\text{TS}] - G_{\text{soln}} (\text{total}) [\text{R}]$$

$$^i G_{\text{soln}} (\text{total}) = G_{\text{soln}} + \Delta\Delta G_{\text{soln}} (\text{thermal})$$

$$^j \Delta G (\text{rxn}) = G_{\text{soln}} (\text{total}) [\text{P}] - G_{\text{soln}} (\text{total}) [\text{R}]$$

$$^k \Delta G (\text{act}) = G_{\text{soln}} (\text{total}) [\text{TS}] - G_{\text{soln}} (\text{total}) [\text{R}]$$

(rxn) stands for reaction and (act) stands for activation.

